Grundlagen der Mathematik 1: Lineare Algebra

Andreas Gathmann

Vorlesungsskript RPTU Kaiserslautern 2023/24

– Version A –

Inhaltsverzeichnis (Grundlagen der Mathematik 1)

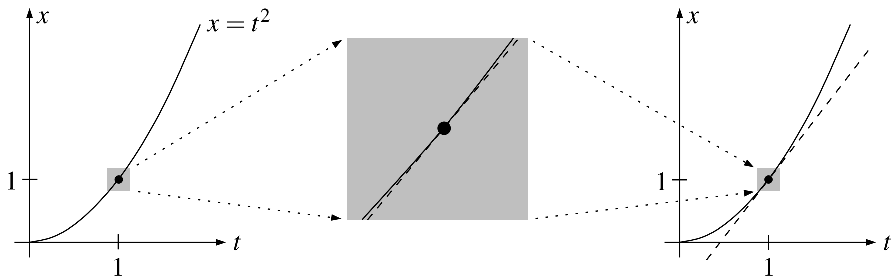
Grundlagen der Mathematik 1: Lineare Algebra

0. Einleitung und Motivation

In diesem Skript – so verspricht es der Titel – wollen wir uns die Grundlagen der Mathematik erarbeiten. Aber was ist das überhaupt, die „Grundlagen der Mathematik“? Bei uns an der TU Kaiserslautern verstehen wir unter dieser Vorlesung die Kombination zweier Themengebiete, die in der Tat das grundlegende Handwerkszeug für nahezu die gesamte Mathematik darstellen, nämlich

- der *Analysis*, d. h. der Untersuchung von Folgen und Grenzwerten, Stetigkeit, sowie der Differential- und Integralrechnung (zunächst in einer und im zweiten Semester dann auch in mehreren Variablen), und
- der *linearen Algebra*, d. h. der Theorie der Vektorräume, linearen Abbildungen und Gleichungssysteme.

Von beiden Gebieten habt ihr ja aus der Schule wahrscheinlich schon eine ungefähre Vorstellung. In der *Analysis* geht es grob gesagt darum, reelle Funktionen *lokal*, also in der Umgebung eines gewählten Punktes, zu untersuchen, und aus diesen Untersuchungen dann wieder Aussagen über die gesamte Funktion zurückzugewinnen. Betrachten wir z. B. ein Auto, das sich entlang einer geraden Strecke bewegt. Der Einfachheit halber nehmen wir an, dass die Position x des Autos nach der Zeit t (in geeigneten Einheiten) durch die Gleichung $x = t^2$ beschrieben werden kann, so dass die Bewegung durch die Kurve im folgenden Bild links dargestellt wird:



Wir wollen diese Bewegung nun nur in einer kleinen Umgebung eines fest gewählten Zeitpunkts, z. B. des Zeitpunkts $t = 1$ (und damit auch $x = 1$) betrachten. Im Bild oben haben wir diese Umgebung grau markiert und in der Mitte stark vergrößert dargestellt. Wir sehen, dass die Kurve in dieser Umgebung fast wie eine *Gerade* aussieht; wir haben diese Gerade gestrichelt eingezeichnet und im Bild rechts auch außerhalb der gewählten Umgebung fortgesetzt. Geometrisch ist diese Gerade natürlich einfach die *Tangente* an die Kurve an der Stelle $t = 1$. Physikalisch repräsentiert die Steigung dieser Geraden die *Geschwindigkeit* des Autos zum betrachteten Zeitpunkt, denn sie gibt ja gerade an, in welchem Verhältnis sich dort die Strecke x mit der Zeit t verändert. Aus der Schule wisst ihr auch schon, wie man diese Steigung ausrechnet: Man muss dazu die gegebene Funktion t^2 *differenzieren* – so dass man die Ableitung $2t$ erhält – und dort die betrachtete Stelle $t = 1$ einsetzen. Die Steigung ist in unserem Fall also gerade $2 \cdot 1 = 2$, und man rechnet sofort nach, dass $x = 2t - 1$ die Gleichung der oben eingezeichneten Tangente ist. Man sagt, dass die Gerade $x = 2t - 1$ eine *lineare Approximation* der ursprünglich gegebenen Funktion $x = t^2$ im Punkt $t = 1$ ist.

Wir können aus der Kenntnis der zurückgelegten Strecke zu jedem Zeitpunkt also die Geschwindigkeit des Autos durch Differenzieren bestimmen. Man kann sich natürlich auch die umgekehrte Frage stellen: Angenommen, der Kilometerzähler eures Autos ist kaputt, aber ihr beobachtet auf eurer Fahrt ständig eure Geschwindigkeit. Könnt ihr dann am Ende der Fahrt trotzdem ausrechnen, wie weit ihr gefahren seid? Dies ist offensichtlich die „Umkehrung“ des Differenzierens – und auch hier wisst ihr aus der Schule natürlich schon, dass dies auf die *Integralrechnung* führen wird.

In der Praxis bewegt man sich mit dem Auto aber nicht immer nur auf einer geraden Strecke, und demzufolge braucht man für die Beschreibung solch einer Bewegung (und natürlich auch vieler anderer natürlich auftretender Prozesse) mehrere Variablen. Wir werden daher auch Abbildungen betrachten, die mehrere Variablen auf mehrere andere abbilden, wie z. B. die folgende Vorschrift, die zwei reelle Zahlen y_1 und y_2 in Abhängigkeit von zwei anderen x_1 und x_2 ausdrückt:

$$
\begin{aligned}
y_1 &= 2x_1 + x_2 \cos x_1 \\
y_2 &= x_1 \, e^{x_2} - x_2
\end{aligned}
$$

Genau wie oben werden wir uns auch hier wieder die Frage stellen, ob wir diese (in diesem Fall recht komplizierte) Funktion in der Nähe eines gegebenen Punktes nicht vielleicht durch eine einfache *lineare* Abhängigkeit annähern können. In der Tat ist dies möglich: Wir werden sehen, dass z. B. in einer kleinen Umgebung des Nullpunkts $(x_1, x_2) = (0,0)$ die obige Vorschrift näherungsweise die gleichen Ergebnisse liefert wie das lineare Gleichungssystem

$$
\begin{aligned}
y_1 &= 2x_1 + x_2 \\
y_2 &= x_1 - x_2.
\end{aligned}
$$

Auch wenn diese Gleichungen natürlich viel einfacher als die ursprünglichen sind, sollte offensichtlich sein, dass auch solche linearen Gleichungssysteme bei wachsender Zahl von Variablen (und in der Praxis sind Hunderte oder Tausende von Variablen keine Seltenheit) recht kompliziert werden können. Wir werden daher einen wesentlichen Teil dieser Vorlesung mit der *linearen Algebra*, also dem Studium derartiger linearer Gleichungssysteme, verbringen. Um dabei überhaupt erst einmal den Notationsaufwand in Grenzen zu halten, tut man dabei gut daran, die Start- und Zielvariablen nicht alle einzeln hinzuschreiben, sondern sie zu sogenannten *Vektoren* zusammenzufassen. In der Tat kann man in dieser Sichtweise die lineare Algebra als das Studium von linearen Abbildungen zwischen Vektorräumen beschreiben.

Wenn wir dann die linearen Abbildungen zwischen mehreren Variablen gut genug verstanden haben, können wir uns im zweiten Teil der Analysis schließlich daran machen, die Differential- und Integralrechnung auf den Fall von mehreren Variablen auszuweiten.

Wir haben damit jetzt relativ kurz umrissen, welcher mathematische Stoff uns in dieser Vorlesung erwartet. Es wird in diesem Skript aber nicht nur darum gehen, mathematische Resultate kennenzulernen. Mindestens ebenso wichtig ist es, das „mathematische Denken" zu lernen, d. h. die Fähigkeit zu entwickeln, mit abstrakten Konzepten umzugehen, exakte logische Schlüsse zu ziehen und Beweise zu führen. Anders als in der Schule oder im Studium z. B. ingenieurswissenschaftlicher Fächer werden wir genau darauf achten, eine „wasserdichte" Theorie aufzubauen: Jeder neue Begriff bzw. jeder neue Satz wird nur unter Verwendung des bisher Bekannten exakt definiert bzw. bewiesen. So sind z. B. Formulierungen in dem Stil „eine Funktion wird durch eine Gerade angenähert" oder „eine Funktion heißt stetig, wenn man sie zeichnen kann, ohne den Stift abzusetzen" als Veranschaulichung unserer Ideen zwar sehr sinnvoll, als exakte mathematische Formulierung jedoch schlichtweg unbrauchbar. Wir werden daher in dieser Vorlesung viel exakter arbeiten als ihr es wahrscheinlich aus der Schule gewohnt seid, und es ist wichtig, dass ihr diese exakte Denkweise verinnerlicht und anzuwenden lernt. Die Mathematik ist ein riesiges Gebäude – viel größer und komplexer als ihr es euch wahrscheinlich im Moment vorstellen könnt – das ständig immer höher gebaut wird, indem schon bewiesene Sätze auf neue Fälle angewendet oder wieder für neue Beweise verwendet werden. Wir starten gerade beim Fundament dieses Gebäudes und können es uns da wirklich nicht leisten herumzupfuschen.

Diese konsequent logische und exakte Herangehensweise ist zwar am Anfang wahrscheinlich ungewohnt, hat jedoch für euch auch einen Vorteil: Etwas überspitzt formuliert erzählen wir euch während des gesamten Studiums eigentlich nur Dinge, die logisch aus dem folgen, was ihr ohnehin schon wusstet. Dadurch ist die Mathematik wahrscheinlich das Studienfach, in dem man am wenigsten auswendig lernen muss – in dem es aber im Gegenzug auch am meisten auf das *Verständnis* des Stoffes ankommt. Je besser euer Verständnis für die Mathematik wird, um so mehr Dinge werden euch letztlich einfach „klar" werden, so dass es euch dann auch viel leichter fällt, sie zu lernen.

Dieses Verständnis für die Mathematik bekommt man aber natürlich nur durch intensiven und *aktiven* Umgang mit dem Stoff, weswegen neben dem Studium der Vorlesung auch die Bearbeitung der Übungsaufgaben besonders wichtig ist.

Heißt das alles nun, dass wir nur mit logischen Argumenten die Mathematik sozusagen „aus dem Nichts“ aufbauen können? Nein, das geht natürlich nicht … von nichts kommt nichts. Man muss am Anfang immer gewisse Dinge als gegeben annehmen, also Aussagen als wahr voraussetzen, die man nicht mehr beweist bzw. beweisen kann, und auf denen dann die gesamte Theorie beruht. Derartige Annahmen bezeichnet man als **Axiome**. Natürlich versucht man in der Mathematik, mit möglichst wenigen und sehr elementaren Axiomen auszukommen, die hoffentlich niemand anzweifeln würde. In der modernen Mathematik ist es üblich, hierfür die grundlegenden Prinzipien der Logik und Mengenlehre zu verwenden (zu denen wir auch gleich in Kapitel 1 Genaueres sagen werden).

In dieser Vorlesung wollen wir uns das Leben allerdings etwas leichter machen und zusätzlich auch die *Existenz und elementaren Eigenschaften der reellen Zahlen* axiomatisch voraussetzen (um welche Eigenschaften es sich hierbei handelt, werden wir natürlich genau angeben). Man kann zwar nur aus den Axiomen der Logik und Mengenlehre beweisen, dass die reellen Zahlen existieren und dass sie die erwarteten Eigenschaften haben, der Beweis wäre zu diesem frühen Zeitpunkt im Studium aber sehr verwirrend und würde euch auch keine großartigen neuen Erkenntnisse bringen. In diesem Sinne starten wir also sozusagen doch nicht ganz beim Fundament unseres „Gebäudes Mathematik“, sondern bereits im ersten Stock.

Im weiteren Verlauf ist dieses Skript dann wie im Inhaltsverzeichnis angegeben in mehrere Teile gegliedert. Diese bauen der Reihe nach aufeinander auf, mit einer Ausnahme: Nach den in jedem Fall benötigten grundlegenden Anfangskapiteln 1 bis 3 sind die Teile „Grundlagen der Mathematik 1: Analysis“ (Kapitel **??** bis **??**) und „Grundlagen der Mathematik 1: Lineare Algebra“ (Kapitel 15 bis 18) *unabhängig voneinander* und können somit in beliebiger Reihenfolge oder auch (wie in den Studienplänen in der Regel vorgesehen) parallel studiert werden.

Aber jetzt genug der Vorrede … beginnen wir nun also unser Studium der Mathematik mit den „Grundlagen der Grundlagen“, den für uns wesentlichen Prinzipien der Logik und Mengenlehre.

1. Etwas Logik und Mengenlehre

Bevor wir mit dem eigentlichen Inhalt der Vorlesung beginnen, müssen wir in diesem Kapitel kurz die exakte mathematische Sprache beschreiben, in der wir unsere Ergebnisse formulieren werden: die der Logik und Mengenlehre. Zentral hierbei sind die Begriffe der *Aussage* (in der Logik) und der *Menge* (in der Mengenlehre).

Da wir es hier mit den ersten beiden Begriffen überhaupt zu tun haben, die in der Mathematik vorkommen, können wir sie natürlich nicht durch bereits bekannte Dinge definieren oder mit bereits bekannten Resultaten ihre Eigenschaften herleiten. Wir müssen sie daher (wie schon in der Einleitung erwähnt) axiomatisch voraussetzen. Wir müssen *voraussetzen*, dass es sinnvoll ist, über logische Aussagen und deren Wahrheit zu reden, dass Mengen überhaupt existieren, dass man Mengen vereinigen und schneiden kann, aus ihnen Elemente auswählen kann, und noch einiges mehr. Wenn ihr euch zum Beispiel auf den Standpunkt stellt, dass ihr nicht an die Existenz von Mengen glaubt, wird euch niemand widerlegen können. Allerdings zweifelt ihr damit dann auch die Existenz der gesamten Mathematik an, wie sie heutzutage betrieben wird – und aus der Tatsache, dass ihr in dieser Vorlesung sitzt, schließe ich einmal, dass das nicht der Fall ist.

Glücklicherweise sind die Dinge, die wir benötigen, jedoch allesamt anschaulich sofort einleuchtend und euch natürlich aus der Schule auch schon hinlänglich bekannt. Ich möchte es euch (und mir) daher ersparen, an dieser Stelle eine vollständige und präzise axiomatische Formulierung der Logik und Mengenlehre hinzuschreiben, zumal das momentan sicher mehr verwirren als helfen würde und außerdem gerade im Bereich der Logik auch zu sehr in die Philosophie abdriften würde. Stattdessen wollen wir uns hier damit begnügen, die für uns wichtigsten Prinzipien und Notationen sowie beliebte Fehlerquellen in verständlicher Sprache zu erklären, auch wenn ein paar Dinge (insbesondere die Begriffsfestlegung – „Definition“ möchte ich es eigentlich gar nicht nennen – einer Aussage und einer Menge) dadurch recht schwammig klingen werden. Außerdem werden wir in Beispielen zur besseren Verdeutlichung bereits hier die reellen Zahlen und ihre einfachsten Eigenschaften (die euch sicherlich bekannt sein werden) benutzen, auch wenn wir diese erst später formalisieren werden. Da es sicher niemanden von euch verwirren wird, werden wir auch die Schreibweise „$x \in \mathbb{R}$“ für „x ist eine reelle Zahl“ schon verwenden, bevor sie in den Notationen 1.13 und 1.15 offiziell eingeführt wird.

1.A Logik

Beginnen wir also mit der Logik. Unter einer **Aussage** verstehen wir (grob gesagt) ein sprachliches Gebilde, das entweder wahr oder falsch ist (wobei wir in der Mathematik natürlich letztlich daran interessiert sind, *wahre* Aussagen herzuleiten). Wichtig sind auch solche sprachlichen Gebilde, in denen freie **Variablen**, also Platzhalter, vorkommen, und die erst beim Einsetzen von Werten für diese Variablen Aussagen liefern. Man bezeichnet sie als **Aussageformen**.

Beispiel 1.1.

(a) $1+1=2$ ist eine wahre, $1+1=3$ eine falsche, und $1+1$ überhaupt keine Aussage.

(b) $x+1=2$ ist eine Aussageform, die beim Einsetzen von $x=1$ in eine wahre, beim Einsetzen jeder anderen reellen Zahl in eine falsche Aussage übergeht.

Bemerkung 1.2. Als Variablen in Aussageformen kann man beliebige Symbole benutzen. Üblich sind neben den normalen lateinischen Klein- und Großbuchstaben auch die griechischen Buchstaben, die wir zur Erinnerung hier auflisten:

$\mathrm{A}\ \alpha$ alpha	$\mathrm{B}\ \beta$ beta	$\Gamma\ \gamma$ gamma	$\Delta\ \delta$ delta	$\mathrm{E}\ \varepsilon$ epsilon	$\mathrm{Z}\ \zeta$ zeta	$\mathrm{H}\ \eta$ eta	$\Theta\ \vartheta$ theta
$\mathrm{I}\ \iota$ iota	$\mathrm{K}\ \kappa$ kappa	$\Lambda\ \lambda$ lambda	$\mathrm{M}\ \mu$ my	$\mathrm{N}\ \nu$ ny	$\Xi\ \xi$ xi	$\mathrm{O}\ o$ omikron	$\Pi\ \pi$ pi
$\mathrm{P}\ \rho$ rho	$\Sigma\ \sigma$ sigma	$\mathrm{T}\ \tau$ tau	$\Upsilon\ \upsilon$ ypsilon	$\Phi\ \varphi$ phi	$\mathrm{X}\ \chi$ chi	$\Psi\ \psi$ psi	$\Omega\ \omega$ omega

Oft verziert man Buchstaben auch noch mit einem Symbol oder versieht sie mit einem Index, um neue Variablen zu erhalten: So sind z. B. $x, x', \tilde{x}, \bar{x}, x_1, x_2, \ldots$ alles Symbole für verschiedene Variablen, die zunächst einmal nichts miteinander zu tun haben (aber tunlichst für irgendwie miteinander zusammenhängende Objekte eingesetzt werden sollten, wenn man den Leser nicht vollends verwirren will).

Notation 1.3 (Zusammengesetzte Aussagen)**.** Sind A und B Aussagen, so lassen sich daraus wie folgt neue bilden:

Symbol	Wahrheitstafel				Bedeutung
A	w	f	w	f	
B	w	w	f	f	
$\neg A$	f	w			nicht A
$A \wedge B$	w	f	f	f	A und B
$A \vee B$	w	w	w	f	A oder B (oder beides): „nicht-ausschließendes Oder“
$A \Leftrightarrow B$	w	f	f	w	A und B sind gleichbedeutend / *äquivalent*, bzw. A genau dann, wenn B
$A \Rightarrow B$	w	w	f	w	aus A folgt B, bzw. wenn A dann B

Die sogenannte **Wahrheitstafel** in den mittleren vier Spalten ist dabei die eigentliche Definition der neuen zusammengesetzten Aussagen. Sie gibt in Abhängigkeit der Wahrheit von A und B (in den ersten beiden Zeilen) an, ob die zusammengesetzte Aussage wahr oder falsch ist.

Bemerkenswert ist hierbei wohl nur die Folgerungsaussage $A \Rightarrow B$, die keine Aussage über die Richtigkeit von A oder B separat macht, sondern nur sagt, dass B wahr ist, wenn auch A es ist. Ist hingegen A falsch, so ist die Folgerungsaussage $A \Rightarrow B$ stets wahr („aus einer falschen Voraussetzung kann man alles folgern“). So ist z. B. $0 = 1 \quad\Rightarrow\quad 2 = 3$ eine wahre Aussage. In der Regel wollen wir uns in der Mathematik aber natürlich mit wahren Aussagen beschäftigen, und neue wahre Aussagen aus alten herleiten. Gerade in Beweisen ist die übliche Verwendung der Notation $A \Rightarrow B$ daher, dass A eine bereits als wahr erkannte Aussage ist, und wir damit nun schließen wollen, dass auch B wahr ist.

Bemerkung 1.4 (Beweise mit Wahrheitstafeln)**.** Wollen wir kompliziertere zusammengesetzte Aussagen miteinander vergleichen, so können wir dies wieder mit Hilfe von Wahrheitstafeln tun. So ist für zwei Aussagen A und B z. B.

$$A \Rightarrow B \quad \text{äquivalent zu} \quad (\neg A) \vee B,$$

denn wenn wir in der Wahrheitstafel

A	w	f	w	f
B	w	w	f	f
$\neg A$	f	w	f	w
$(\neg A) \vee B$	w	w	f	w

mit Hilfe der Definitionen von $\neg$ und $\vee$ aus Notation 1.3 zunächst $\neg A$ und dann $(\neg A) \vee B$ berechnen, sehen wir, dass das Ergebnis mit $A \Rightarrow B$ übereinstimmt. Nach der Bemerkung aus Notation 1.3 ist dies auch anschaulich klar: Die Folgerungsaussage $A \Rightarrow B$ ist ja genau dann wahr, wenn A falsch (also $\neg A$ wahr) ist, oder wenn B wahr ist (oder beides).

Genauso zeigt man die ebenfalls einleuchtende Aussage, dass

$$A \Leftrightarrow B \quad \text{äquivalent zu} \quad (A \Rightarrow B) \wedge (B \Rightarrow A)$$

ist – was auch die übliche Art ist, wie man eine Äquivalenz zeigt: Man zeigt separat die beiden Folgerungen $A \Rightarrow B$ und $B \Rightarrow A$.

Notation 1.5. Folgerungen („$\Rightarrow$") und Äquivalenzen („$\Leftrightarrow$") sind natürlich zwei verschiedene Dinge, die man nicht durcheinanderwerfen darf (auch wenn das in der Schule wahrscheinlich manchmal nicht so genau genommen wird). Es hat sich jedoch in der Mathematik eingebürgert, bei *Definitionen* von Begriffen durch eine äquivalente, definierende Eigenschaft die Sprechweise „wenn" anstatt des eigentlich korrekten „genau dann, wenn" zu verwenden: So würde man z. B. als Definition des Begriffs einer geraden Zahl hinschreiben

> „Eine ganze Zahl x heißt gerade, wenn $\frac{x}{2}$ eine ganze Zahl ist",

obwohl man genau genommen natürlich meint

> „Eine ganze Zahl x heißt *genau dann* gerade, wenn $\frac{x}{2}$ eine ganze Zahl ist".

Eine gewöhnliche Folgerungsaussage wie z. B. die wahre Aussage

> „Wenn eine ganze Zahl x positiv ist, dann ist auch $x+1$ positiv"

ist dagegen immer nur in einer Richtung zu verstehen; hier wird also nicht behauptet, dass mit $x+1$ auch x immer positiv sein muss (was ja auch falsch wäre).

Notation 1.6 (Quantoren)**.** Ist A eine Aussageform, in der eine freie Variable x vorkommt – wir schreiben dies dann auch als $A(x)$ – so setzen wir

Symbol	Bedeutung
$\forall x : A(x)$	für alle x gilt $A(x)$
$\exists x : A(x)$	es gibt ein x mit $A(x)$

Die beiden Symbole $\forall$ und $\exists$ bezeichnet man als **Quantoren**. Beachte, dass diese beiden Quantoren *nicht* miteinander vertauschbar sind: So besagt z. B. die Aussage

$$\forall x \in \mathbb{R}\ \exists y \in \mathbb{R} : y > x$$

„zu jeder reellen Zahl x gibt es eine Zahl y, die größer ist" (was offensichtlich wahr ist), während die Umkehrung der beiden Quantoren die Aussage

$$\exists y \in \mathbb{R}\ \forall x \in \mathbb{R} : y > x$$

„es gibt eine reelle Zahl y, die größer als jede reelle Zahl x ist" liefern würde (was ebenso offensichtlich falsch ist). Der Unterschied besteht einfach darin, dass im ersten Fall zuerst das x gewählt werden muss und dann ein y dazu existieren muss (das von x abhängen darf), während es im zweiten Fall *dasselbe* y für alle x tun müsste.

Bemerkung 1.7. Jede Aussage lässt sich natürlich auf viele Arten aufschreiben, sowohl als deutscher Satz als auch als mathematische Formel. Die gerade eben betrachtete Aussage könnte man z. B. auf die folgenden (absolut gleichwertigen) Arten aufschreiben:

(a) $\forall x \in \mathbb{R}\ \exists y \in \mathbb{R} : y > x$.

(b) Es sei $x \in \mathbb{R}$. Dann gibt es ein $y \in \mathbb{R}$ mit $y > x$.

(c) Zu jeder reellen Zahl gibt es noch eine größere.

Welche Variante man beim Aufschreiben wählt ist weitestgehend Geschmackssache. Die Formulierung einer Aussage als deutscher Satz hat den Vorteil, dass wir sie oft leichter verstehen können, weil wir die deutsche Sprache schon länger kennen als die mathematische. Wenn wir uns jedoch erst einmal an die mathematische Sprache gewöhnt haben, wird auch sie ihre Vorzüge bekommen: Sie ist deutlich kürzer und besser logisch strukturiert. Wir werden im Folgenden beide Schreibweisen mischen und jeweils diejenige wählen, mit der unsere Aussagen (hoffentlich) am einfachsten verständlich werden.

Wenn wir mathematische Symbole verwenden, müssen wir diese aber auch stets in ihrer korrekten Notation und nicht als „Abkürzungen" für deutsche Wörter verwenden: Man würde die Aussage

„2 und 4 sind gerade Zahlen“ sicher niemals schreiben als „$2 \wedge 4$ sind gerade Zahlen“, und analog genauso wenig „Es gilt $n \geq 0$ für alle $n \in \mathbb{N}$“ als „Es gilt $n \geq 0 \, \forall n \in \mathbb{N}$“.

Bemerkung 1.8 (Negationen)**.** Es ist wichtig zu wissen, wie man von einer Aussage das Gegenteil, also die „Verneinung“ bildet. Da hierbei oft Fehler gemacht werden, wollen wir die allgemeinen Regeln hierfür kurz auflisten:

(a) $\neg(\neg A) \Leftrightarrow A$: Ist es falsch, dass A falsch ist, so bedeutet dies genau, dass A wahr ist.

(b) $\neg(A \wedge B) \Leftrightarrow (\neg A) \vee (\neg B)$: Das Gegenteil von „$A$ und B sind richtig“ ist „A oder B ist falsch“.

(c) $\neg(A \vee B) \Leftrightarrow (\neg A) \wedge (\neg B)$: Das Gegenteil von „$A$ oder B ist richtig“ ist „A und B sind falsch“.

(d) $\neg(\forall x : A(x)) \Leftrightarrow \exists x : \neg A(x)$: Das Gegenteil von „für alle x gilt $A(x)$“ ist „es gibt ein x, für das $A(x)$ falsch ist“.

(e) $\neg(\exists x : A(x)) \Leftrightarrow \forall x : \neg A(x)$: Das Gegenteil von „es gibt ein x, für das $A(x)$ gilt“ ist „für alle x ist $A(x)$ falsch“.

Man kann also sagen, dass eine Verneinung dazu führt, dass „und“ mit „oder“ sowie „für alle“ mit „es gibt“ vertauscht wird. So ist z. B. das Gegenteil der Aussage

„In Frankfurt haben *alle* Haushalte Strom *und* fließendes Wasser“

die Aussage

„In Frankfurt *gibt es* einen Haushalt, der keinen Strom *oder* kein fließendes Wasser hat“.

Beispiel 1.9 (Negation einer Folgerung)**.** Wollen wir eine Folgerung $A \Rightarrow B$ verneinen, so können wir sie zunächst mit Bemerkung 1.4 zu $(\neg A) \vee B$ umformen, und erhalten nach Bemerkung 1.8 als Umkehrung dann $A \wedge \neg B$. Dies ist auch anschaulich einleuchtend: Die Folgerungsaussage „wenn A dann B“ ist genau dann falsch, wenn die Voraussetzung A zwar gilt, die Behauptung B aber nicht. Wir sehen also:

$$\text{Die Verneinung einer Folgerung } \; A \Rightarrow B \; \text{ ist } \; A \wedge \neg B.$$

Bemerkung 1.10 (Widerspruchsbeweis)**.** Eine oft vorkommende Anwendung der Regeln für die Verneinung von Aussagen ist der sogenannte **Widerspruchsbeweis** bzw. Beweis durch **Kontraposition**. Nach Bemerkung 1.4 gesehen ist die Folgerung $A \Rightarrow B$ („aus A folgt B“) gleichbedeutend mit $(\neg A) \vee B$. Damit ist diese Aussage nach Bemerkung 1.8 (a) auch äquivalent zu $(\neg(\neg B)) \vee (\neg A)$, also zu $\neg B \Rightarrow \neg A$. Mit anderen Worten: Haben wir eine Schlussfolgerung $A \Rightarrow B$ zu beweisen, so können wir genauso gut $(\neg B) \Rightarrow (\neg A)$ zeigen, d. h. *wir können annehmen, dass die zu zeigende Aussage B falsch ist und dies dann zu einem Widerspruch führen bzw. zeigen, dass dann auch die Voraussetzung A falsch sein muss.*

Beispiel 1.11. Hier sind zwei Beispiele für die Anwendung der Prinzipien aus Bemerkung 1.8 und 1.10 – und auch unsere ersten Beispiele dafür, wie man Beweise von Aussagen exakt aufschreiben kann.

(a) Einen Beweis durch Widerspruch könnte man z. B. so aufschreiben:

Behauptung: Für alle $x \in \mathbb{R}$ gilt $2x+1 > 0$ oder $2x-1 < 0$.

Beweis: Angenommen, die Behauptung wäre falsch, d. h. (nach Bemerkung 1.8 (c) und (d)) es gäbe ein $x \in \mathbb{R}$ mit

$$2x+1 \leq 0 \quad (1) \qquad \text{und} \qquad 2x-1 \geq 0 \quad (2).$$

Für dieses x würde dann folgen, dass

$$0 \overset{(1)}{\geq} 2x+1 = 2x-1+2 \overset{(2)}{\geq} 0+2 = 2.$$

Dies ist aber ein Widerspruch. Also war unsere Annahme falsch und somit die zu beweisende Aussage richtig. □

Das dabei verwendete Symbol „$\square$“ ist die übliche Art, das Ende eines Beweises zu kennzeichnen. Zur Verdeutlichung haben wir die beiden Ungleichungen mit (1) und (2) markiert, um später angeben zu können, wo sie verwendet werden.

(b) Manchmal weiß man von einer Aussage aufgrund der Aufgabenstellung zunächst einmal noch nicht, ob sie wahr oder falsch ist. In diesem Fall muss man sich dies natürlich zuerst überlegen – und, falls die Aussage falsch ist, ihre Negation beweisen. Als Beispiel dafür betrachten wir die Aufgabe

Man beweise oder widerlege: Für alle $x \in \mathbb{R}$ gilt $2x+1<0$ oder $2x-1>0$.

In diesem Fall merkt man schnell, dass die Aussage falsch sein muss, weil die Ungleichungen schon für den Fall $x=0$ nicht stimmen. Man könnte als Lösung der Aufgabe unter Beachtung der Negationsregeln aus Bemerkung 1.8 also aufschreiben:

Behauptung: Die Aussage ist falsch, d. h. es gibt ein $x \in \mathbb{R}$ mit $2x+1 \geq 0$ und $2x-1 \leq 0$.

Beweis: Für $x=0$ ist $2x+1=1 \geq 0$ und $2x-1=-1 \leq 0$. $\square$

Beachte, dass dies ein vollständiger Beweis ist: *Um eine allgemeine Aussage zu widerlegen, genügt es, ein Gegenbeispiel dafür anzugeben.*

Bemerkung 1.12. Bevor wir unsere kurze Auflistung der für uns wichtigen Prinzipien der Logik beenden, wollen wir noch kurz auf ein paar generelle Dinge eingehen, die man beim Aufschreiben mathematischer Beweise oder Rechnungen beachten muss.

Dass wir bei unseren logischen Argumenten sauber und exakt arbeiten – also z. B. nicht Folgerungen, die keine Äquivalenzen sind, in der falschen Richtung verwenden, „für alle“ mit „es gibt“ verwechseln oder ähnliches – sollte sich von selbst verstehen. Die folgende kleine Geschichte hilft vielleicht zu verstehen, was damit gemeint ist.

Ein Ingenieur, ein Physiker und ein Mathematiker fahren mit dem Zug nach Frankreich und sehen dort aus dem Fenster des Zuges ein schwarzes Schaf.

Da sagt der Ingenieur: „Oh, in Frankreich sind die Schafe schwarz!“

Darauf der Physiker: „Nein ... wir wissen jetzt nur, dass es in Frankreich mindestens ein schwarzes Schaf gibt.“

Der Mathematiker: „Nein ... wir wissen nur, dass es in Frankreich mindestens ein Schaf gibt, das auf mindestens einer Seite schwarz ist.“

Es gibt aber noch einen weiteren sehr wichtigen Punkt, der selbst von fortgeschritteneren Studenten leider oft nicht beachtet wird: In der Regel werden wir beim Aufschreiben sowohl Aussagen notieren wollen, die wir erst noch zeigen wollen (um schon einmal zu sagen, worauf wir hinaus wollen), als auch solche, von denen wir bereits wissen, dass sie wahr sind (z. B. weil sie für die zu zeigende Behauptung als wahr vorausgesetzt werden oder weil sie sich logisch aus irgendetwas bereits Bekanntem ergeben haben). Es sollte offensichtlich sein, dass wir Aussagen mit derartig verschiedenen Bedeutungen für die Argumentationsstruktur nicht einfach kommentarlos nebeneinander schreiben dürfen, wenn noch jemand in der Lage sein soll, die Argumente nachzuvollziehen. Betrachten wir z. B. noch einmal unseren Beweis aus Beispiel 1.11 (a) oben, so wäre eine Art des Aufschreibens in folgendem Stil (wie man es leider oft sieht)

$$\begin{array}{lll} 2x+1>0 & \text{oder} & 2x-1<0 \\ 2x+1 \leq 0 & & 2x-1 \geq 0 \\ \multicolumn{3}{l}{0 \geq 2x+1 = 2x-1+2 \geq 0+2 = 2} \end{array}$$

völlig inakzeptabel, obwohl hier natürlich letztlich die gleichen Aussagen stehen wie oben. Kurz gesagt:

Von *jeder* aufgeschriebenen Aussage muss für den Leser *sofort* und *ohne eigenes Nachdenken* ersichtlich sein, welche Rolle sie in der Argumentationsstruktur spielt: Ist es z. B. eine noch zu zeigende Behauptung, eine Annahme oder eine Folgerung (und wenn ja, aus was)?

Dies bedeutet allerdings nicht, dass wir ganze Aufsätze schreiben müssen. Eine (schon recht platzoptimierte) Art, den Beweis aus Beispiel 1.11 (a) aufzuschreiben, wäre z. B.

Angenommen, es gäbe ein $x \in \mathbb{R}$ mit $2x+1 \leq 0$ und $2x-1 \geq 0$.
Dann wäre $0 \geq 2x+1 = 2x-1+2 \geq 0+2 = 2$, Widerspruch. $\square$

1.B Mengenlehre

Nachdem wir die wichtigsten Regeln der Logik behandelt haben, wenden wir uns jetzt der Mengenlehre zu. Die gesamte moderne Mathematik basiert auf diesem Begriff der Menge, der ja auch schon aus der Schule hinlänglich bekannt ist. Zur Beschreibung, was eine Menge ist, zitiert man üblicherweise die folgende Charakterisierung von Georg Cantor (1845–1918):

„Eine **Menge** ist eine Zusammenfassung von bestimmten, wohlunterschiedenen Objekten unserer Anschauung oder unseres Denkens zu einem Ganzen.“

Die in einer Menge zusammengefassten Objekte bezeichnet man als ihre **Elemente**.

Notation 1.13.

(a) Wir schreiben $x \in M$, falls x ein Element der Menge M ist, und $x \notin M$ andernfalls.

(b) Die einfachste Art, eine Menge konkret anzugeben, besteht darin, ihre Elemente in geschweiften Klammern aufzulisten, wobei es auf die Reihenfolge und Mehrfachnennungen nicht ankommt. So sind z. B. $\{1,2,3\}$ und $\{2,3,3,1\}$ zwei Schreibweisen für dieselbe Menge. Man beachte, dass die Elemente einer Menge nicht unbedingt Zahlen sein müssen – wir könnten z. B. auch die Menge aller Studenten in dieser Vorlesung betrachten. In der Tat kommen in der Mathematik sogar oft Mengen vor, deren Elemente selbst wieder Mengen sind; so ist z. B. $\{\{1,2,3\},\{1,2,3,4\}\}$ eine gültige Menge (mit 2 Elementen).

(c) Man kann die Elemente einer Menge auch durch eine beschreibende Eigenschaft angeben: $\{x : A(x)\}$ bezeichnet die Menge aller Objekte x, für die die Aussage $A(x)$ wahr ist, wie z. B. in $\{x \in \mathbb{R} : x^2 = 1\} = \{-1,1\}$.

(d) Die Menge $\{\}$ ohne Elemente, die sogenannte **leere Menge**, bezeichnen wir mit $\emptyset$.

(e) Man schreibt $M \subset N$ (oder $N \supset M$), wenn jedes Element von M auch Element von N ist, bzw. in der Quantorenschreibweise von Notation 1.6 wenn

$$\forall x : x \in M \Rightarrow x \in N.$$

In diesem Fall sagt man, dass M eine **Teilmenge** von N (oder N eine **Obermenge** von M) ist. Beachte, dass M und N dabei auch gleich sein können; in der Tat ist offensichtlich

$$M = N \quad \text{genau dann, wenn} \quad M \subset N \text{ und } N \subset M.$$

Oft wird man eine Gleichheit $M = N$ von Mengen auch so beweisen, dass man separat $M \subset N$ und $N \subset M$ zeigt.

Wenn wir ausdrücken wollen, dass M eine Teilmenge von N und nicht gleich N ist, so schreiben wir dies als $M \subsetneq N$ und sagen, dass M eine **echte Teilmenge** von N ist. Es ist wichtig, dies von der Aussage $M \not\subset N$ zu unterscheiden, die bedeutet, dass M keine Teilmenge von N ist.

Achtung: Manchmal wird in der Literatur das Symbol „$\subset$“ für *echte* Teilmengen und „$\subseteq$“ für nicht notwendig echte Teilmengen verwendet.

(f) Hat eine Menge M nur endlich viele Elemente, so nennt man M eine **endliche Menge** und schreibt die Anzahl ihrer Elemente als $|M|$. Andernfalls setzt man formal $|M| = \infty$.

Bemerkung 1.14 (Russellsches Paradoxon)**.** Die oben gegebene Charakterisierung von Mengen von Cantor ist aus mathematischer Sicht natürlich sehr schwammig. In der Tat hat Bertrand Russell kurz darauf bemerkt, dass sie sogar schnell zu Widersprüchen führt. Er betrachtet dazu

$$M = \{A : A \text{ ist eine Menge mit } A \notin A\}, \tag{$*$}$$

also „die Menge aller Mengen, die sich nicht selbst als Element enthalten". Sicherlich ist es eine merkwürdige Vorstellung, dass eine Menge sich selbst als Element enthalten könnte – im Sinne von Cantors Charakterisierung wäre die Definition $(*)$ aber zulässig. Fragen wir uns nun allerdings, ob sich die so konstruierte Menge M selbst als Element enthält, so erhalten wir sofort einen Widerspruch: Wenn $M \in M$ gilt, so würde das nach der Definition $(*)$ ja gerade bedeuten, dass $M \notin M$ ist – und das wiederum, dass doch $M \in M$ ist. Man bezeichnet dies als das *Russellsche Paradoxon*.

Die Ursache für diesen Widerspruch ist, dass die Definition $(*)$ rückbezüglich ist: Wir wollen eine neue Menge M konstruieren, verwenden dabei aber auf der rechten Seite der Definition *alle Mengen*, also u. a. auch die Menge M, die wir gerade erst definieren wollen. Das ist in etwa so, als würdet ihr im Beweis eines Satzes die Aussage des Satzes selbst verwenden – und das ist natürlich nicht zulässig.

Man muss bei der Festlegung, was Mengen sind und wie man sie bilden kann, also eigentlich viel genauer vorgehen, als es Cantor getan hat. Heutzutage verwendet man hierzu in der Regel das im Jahre 1930 aufgestellte Axiomensystem von Zermelo und Fraenkel, das genau angibt, wie man aus bekannten Mengen neue konstruieren darf: z. B. indem man sie schneidet oder vereinigt, oder aus bereits bekannten Mengen Elemente mit einer bestimmten Eigenschaft auswählt. Wir wollen dies hier in dieser Vorlesung aber nicht weiter thematisieren und uns mit der naiven Mengencharakterisierung von Cantor begnügen (sowie der Versicherung meinerseits, dass schon alles in Ordnung ist, wenn wir neue Mengen immer nur aus alten konstruieren und keine rückbezüglichen Definitionen hinschreiben). Genaueres zum Zermelo-Fraenkel-Axiomensystem könnt ihr z. B. in [E, Kapitel 13] nachlesen.

Notation 1.15 (Reelle Zahlen)**.** Unser wichtigstes Beispiel für eine Menge ist die Menge der **reellen Zahlen**, die wir mit $\mathbb{R}$ bezeichnen werden. Wir wollen die Existenz der reellen Zahlen in dieser Vorlesung axiomatisch voraussetzen und begnügen uns daher an dieser Stelle damit zu sagen, dass man sie sich als die Menge der Punkte auf einer Geraden (der „Zahlengeraden") vorstellen kann. Zusätzlich werden wir in den nächsten beiden Kapiteln die mathematischen Eigenschaften von $\mathbb{R}$ exakt angeben (und ebenfalls axiomatisch voraussetzen) – und zwar genügend viele Eigenschaften, um $\mathbb{R}$ dadurch eindeutig zu charakterisieren.

Ich möchte hier noch einmal betonen, dass man die Existenz und die Eigenschaften der reellen Zahlen eigentlich nicht voraussetzen müsste: Man kann das auch allein aus den Axiomen der Logik und Mengenlehre herleiten! Dies wäre jedoch relativ aufwändig und würde euch im Moment mehr verwirren als helfen, daher wollen wir hier darauf verzichten. Wer sich trotzdem dafür interessiert, kann die Einzelheiten hierzu in [E, Kapitel 1 und 2] nachlesen.

Außer den reellen Zahlen sind vor allem noch die folgenden Teilmengen von $\mathbb{R}$ wichtig:

(a) die Menge $\mathbb{N} = \{0, 1, 2, \dots\}$ der **natürlichen Zahlen** (Achtung: Es gibt Bücher, in denen die 0 nicht mit zu den natürlichen Zahlen gezählt wird!);

(b) die Menge $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ der **ganzen Zahlen**;

(c) die Menge $\mathbb{Q} = \{\frac{p}{q} : p, q \in \mathbb{Z}, q \neq 0\}$ der **rationalen Zahlen**.

Offensichtlich sind diese Mengen ineinander enthalten: Es gilt $\mathbb{N} \subsetneq \mathbb{Z} \subsetneq \mathbb{Q} \subsetneq \mathbb{R}$. Teilmengen von $\mathbb{R}$, die durch Ungleichungen gegeben sind, schreiben wir in der Regel, indem wir die Ungleichungsbedingung als Index an das Symbol $\mathbb{R}$ schreiben, z. B. $\mathbb{R}_{\geq 0}$ für die Menge $\{x \in \mathbb{R} : x \geq 0\}$ aller nicht-negativen Zahlen.

Definition 1.16. Sind M und N Mengen, so bezeichnen wir mit ...

(a) $M \cap N := \{x : x \in M \text{ und } x \in N\}$ die **Schnittmenge** von M und N. Gilt $M \cap N = \emptyset$, so sagen wir, dass M und N **disjunkt** sind.

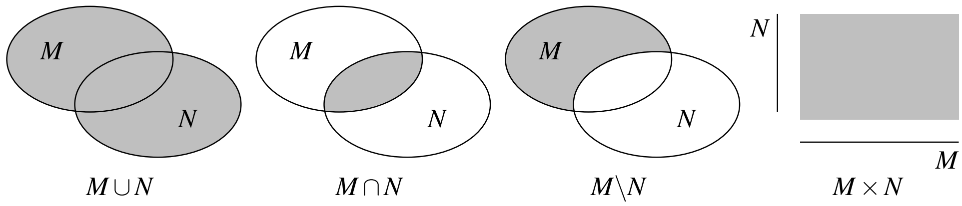
(b) $M \cup N := \{x : x \in M \text{ oder } x \in N\}$ die **Vereinigungsmenge** von M und N. im Fall einer **disjunkten Vereinigung** mit $M \cap N = \emptyset$ schreiben wir statt $M \cup N$ auch $M \,\dot\cup\, N$.

(c) $M \setminus N := \{x : x \in M \text{ und } x \notin N\}$ die **Differenzmenge** von M und N.

(d) $M \times N := \{(x,y) : x \in M, y \in N\}$ die **Produktmenge** bzw. das Produkt von M und N. Die Schreibweise (x,y) steht hierbei für ein **geordnetes Paar**, d. h. einfach für die Angabe eines Elements aus M und eines aus N (wobei es auch im Fall $M = N$ auf die Reihenfolge ankommt, d. h. (x,y) ist genau dann gleich (x',y') wenn $x = x'$ und $y = y'$). Im Fall $M = N$ schreibt man $M \times N = M \times M$ auch als M^2.

(e) $\mathscr{P}(M) := \{A : A \text{ ist Teilmenge von } M\}$ die **Potenzmenge** von M.

Das Symbol „:=“ bedeutet hierbei, dass der Ausdruck auf der linken Seite durch die rechte Seite definiert wird. Die Konstruktionen (a) bis (d) können durch die folgenden Bilder veranschaulicht werden. Natürlich sind sie auch für mehr als zwei Mengen möglich; aus der Schule kennt ihr zum Beispiel sicher den Fall $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$.



Die Potenzmenge $\mathscr{P}(M)$ aller Teilmengen einer gegebenen Menge M lässt sich dagegen nicht so einfach durch ein Bild darstellen. Es ist z. B.

$$\mathscr{P}(\{0,1\}) = \{\emptyset, \{0\}, \{1\}, \{0,1\}\}.$$

Aufgabe 1.17. Wie lautet die Negation der folgenden Aussagen? Formuliere außerdem die Aussage (a) in Worten (also analog zu (b)) sowie die Aussage (b) mit Quantoren und anderen mathematischen Symbolen (also analog zu (a)).

(a) $\forall n \in \mathbb{N}\ \exists m \in \mathbb{N} : n = 2m$.

(b) Zwischen je zwei verschiedenen reellen Zahlen gibt es noch eine weitere reelle Zahl.

(c) Sind M,N,R Mengen mit $R \subset N \subset M$, so ist $M \setminus N \subset M \setminus R$.

Aufgabe 1.18. Es seien A, B, C Aussagen und M, N, R Mengen. Man zeige:

(a) $A \vee (B \wedge C) \Leftrightarrow (A \vee B) \wedge (A \vee C)$ und $A \wedge (B \vee C) \Leftrightarrow (A \wedge B) \vee (A \wedge C)$.

(b) $M \cup (N \cap R) = (M \cup N) \cap (M \cup R)$ und $M \cap (N \cup R) = (M \cap N) \cup (M \cap R)$.

Aufgabe 1.19. Welche der folgenden Aussagen sind für beliebige gegebene x und M äquivalent zueinander? Zeige jeweils die Äquivalenz bzw. widerlege sie durch ein Gegenbeispiel.

(a) $x \in M$	(b) $\{x\} \subset M$	(c) $\{x\} \cap M \neq \emptyset$
(d) $\{x\} \in M$	(e) $\{x\} \setminus M = \emptyset$	(f) $M \setminus \{x\} = \emptyset$

Aufgabe 1.20. Man beweise oder widerlege: Für alle Mengen $A \subset M$ und $A' \subset M'$ gibt es Teilmengen B und C von M sowie B' und C' von M', so dass

$$(M \times M') \setminus (A \times A') = (B \times B') \,\dot\cup\, (C \times C').$$

Könnt ihr die Aussage durch eine Skizze veranschaulichen?

2. Relationen und Funktionen

Nachdem wir Mengen eingeführt haben, wollen wir nun auch mehrere von ihnen miteinander in Beziehung setzen können. Das Grundkonzept hierfür ist das einer Relation.

Definition 2.1 (Relationen)**.** Es seien M und N zwei Mengen. Eine **Relation** zwischen M und N ist eine Teilmenge R des Produkts $M \times N$. Für $x \in M$ und $y \in N$ mit $(x,y) \in R$ sagen wir dann „*x steht (bezüglich R) in Relation zu y*". Ist $M = N$, so nennen wir R auch eine *Relation auf M*.

Bemerkung 2.2. Um eine Relation R anzugeben, also eine Teilmenge $R \subset M \times N$ zu definieren, müssen wir demzufolge einfach für alle Paare (x,y) mit $x \in M$ und $y \in N$ festlegen, ob $(x,y) \in R$ gelten, also ob x in Relation zu y stehen soll.

Wie wir in diesem Kapitel sehen werden, werden Relationen in der Mathematik für völlig unterschiedliche Konzepte verwendet – z. B. um Zahlen miteinander zu vergleichen wie in Beispiel 2.3 (b), um eine Menge auf eine andere abzubilden wie in Abschnitt 2.A, oder um die Elemente einer Menge nach bestimmten Kriterien zu Klassen zusammenzufassen wie in Abschnitt 2.B. Dementsprechend sind für die Aussage „x steht bezüglich R in Relation zu y" auch je nach Anwendung ganz unterschiedliche Notationen üblich. Für allgemeine, nicht näher spezifizierte Relationen schreibt man hierfür oft $x R y$.

Beispiel 2.3.

(a) Es sei M die Menge aller derzeit an der TU Kaiserslautern eingeschriebenen Studenten und N die Menge aller in diesem Semester angebotenen Vorlesungen. Dann ist

$$R = \{(x,y) : x \text{ hört } y\}$$

eine Relation zwischen M und N.

(b) Für $M = N = \mathbb{R}$ betrachten wir die Relation

$$R = \{(x,y) : x,y \in \mathbb{R} \text{ mit } x < y\},$$

für die x also genau dann in Relation zu y steht, wenn $x < y$ gilt. Man nennt R deshalb auch die *Kleiner-Relation* auf $\mathbb{R}$. Die Notation „$x R y$" aus Bemerkung 2.2 stimmt in diesem Fall also mit der Schreibweise „$x < y$" überein, wenn man die Relation R direkt mit dem Symbol „$<$" bezeichnet. In der Tat ist es aus diesem Grund bei manchen Relationen üblich, sie gleich mit Symbolen statt mit Buchstaben zu benennen.

2.A Funktionen

Die mit Abstand wichtigsten Relationen sind ohne Zweifel die Funktionen, die ihr natürlich bereits hinlänglich aus der Schule kennt. Wir wollen sie hier nun exakt einführen und ihre ersten Eigenschaften untersuchen.

Definition 2.4 (Funktionen)**.** Es seien M und N zwei Mengen.

(a) Eine **Funktion** oder **Abbildung** f von M nach N, geschrieben $f : M \to N$, ist eine Relation zwischen M und N, bezüglich der jedes Element x von M zu *genau einem* Element y von N in Relation steht. Wir schreiben dies dann als $x \mapsto y$ oder $y = f(x)$ und sagen, y ist das **Bild** von x unter f bzw. der **Wert** von f in x.

(b) Für eine Funktion $f : M \to N$ bezeichnet man die Menge M als **Definitionsmenge**, **Startmenge** oder **Startraum** von f. Die Menge N heißt **Zielmenge** oder **Zielraum** von f.

Bemerkung 2.5.

(a) Um eine Funktion komplett festzulegen, müssen wir zuerst einmal den Start- und Zielraum angeben, und dann schließlich noch von jedem Element des Startraums sagen, auf welches Element des Zielraums es abgebildet wird. In welcher Form wir diese Zuordnung angeben – ob durch eine Formel, durch explizites Auflisten der Funktionswerte aller Elemente des Startraums, oder irgendwie anders – spielt dabei keine Rolle. So sind z. B.

$$f\colon \{0,1\} \to \mathbb{R},\, x \mapsto 2x^2, \quad g\colon \{0,1\} \to \mathbb{R},\, x \mapsto 2x^3, \quad h\colon \{0,1\} \to \mathbb{R},\, x \mapsto \begin{cases} 0 & \text{für } x = 0, \\ 2 & \text{für } x = 1 \end{cases}$$

trotz ihrer ganz verschieden aussehenden Vorschriften dieselbe Funktion, da alle drei den gleichen Start- und Zielraum haben und aus den gleichen Zuordnungen $0 \mapsto 0$ und $1 \mapsto 2$ bestehen. Mit anderen Worten sind zwei Funktionen $f, g\colon M \to N$ also genau dann gleich, wenn sie an jedem Punkt die gleichen Werte besitzen, also wenn gilt

$$\forall x \in M : f(x) = g(x).$$

(b) Man sieht leider oft, dass eine Funktion $f\colon M \to N$ als $f(x)$ geschrieben wird. Es ist wichtig zu verstehen, dass diese Notation gemäß Definition 2.4 falsch ist: Mit $f(x)$ wird *der Wert der Funktion f in einem Punkt $x \in M$* bezeichnet. Somit ist $f(x)$ (für gegebenes x) ein Element von N, und damit ein ganz anderes mathematisches Objekt als die Funktion selbst, die wir nur mit f bezeichnen und die eine Relation zwischen M und N ist. Dies mag auf den ersten Blick spitzfindig erscheinen – wir werden aber später noch oft Mengen sehen, deren Elemente Funktionen sind, und dann ist es natürlich wichtig, dies von der Menge ihrer Funktions*werte* zu unterscheiden.

Beispiel 2.6.

(a) Die Zuordnungen

$$f\colon \mathbb{R} \to \mathbb{R},\, x \mapsto \frac{1}{x} \qquad \text{und} \qquad g\colon \mathbb{R} \to \mathbb{R},\, x \mapsto \begin{cases} x+1 & \text{für } x \le 0 \\ x & \text{für } x \ge 0 \end{cases}$$

sind in dieser Form keine zulässigen Funktionsdefinitionen, weil im Fall von f der Zahl 0 kein gültiger Funktionswert zugeordnet wird und im Fall g für die Zahl 0 zwei (sich widersprechende) Festlegungen des Funktionswertes gemacht werden. Dies lässt sich jedoch in beiden Fällen leicht reparieren, z. B. indem man die Festlegungen abändert in

$$f\colon \mathbb{R}\backslash\{0\} \to \mathbb{R},\, x \mapsto \frac{1}{x} \qquad \text{und} \qquad g\colon \mathbb{R} \to \mathbb{R},\, x \mapsto \begin{cases} x+1 & \text{für } x < 0 \\ x & \text{für } x \ge 0 \end{cases}.$$

(b) Zu jeder Menge M gibt es die **identische Abbildung**

$$\mathrm{id}_M\colon M \to M,\, x \mapsto x,$$

die jedes Element auf sich selbst abbildet.

(c) Ist $f\colon M \to N$ eine Abbildung und $A \subset M$ eine Teilmenge des Startraums, so erhält man durch die Einschränkung der Definitionsmenge von M auf A eine neue Abbildung, die wir mit

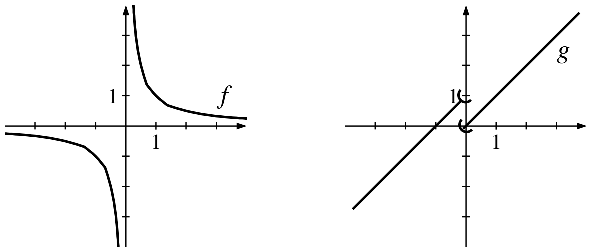
$$f|_A\colon A \to N,\, x \mapsto f(x)$$

bezeichnen und die die **Einschränkung** von f auf A genannt wird. Genauso kann man natürlich auch die Zielmenge N auf eine Teilmenge B einschränken, wenn f nur Werte in B annimmt. Es ist üblich, bei einer derartigen Einschränkung der Zielmenge immer noch den gleichen Namen für die Abbildung zu verwenden, also dann $f\colon M \to B$ zu schreiben (auch wenn es sich dabei um eine andere Funktion als das ursprüngliche $f\colon M \to N$ handelt).

Bemerkung 2.7 (Graph einer Abbildung)**.** Zu einer Abbildung $f\colon M \to N$ heißt die Menge

$$\{(x, f(x)) : x \in M\} \quad \subset M \times N$$

der **Graph** von f. Sind M und N Teilmengen von $\mathbb{R}$, so ist dieser Graph also eine Teilmenge von $\mathbb{R}^2$, und man kann ihn leicht zeichnen und dadurch die Abbildung veranschaulichen. Für die Abbildungen aus Beispiel 2.6 (a) sieht dies z. B. so aus:



Beachte, dass dieser Graph nach den Definitionen 2.1 und 2.4 eigentlich sogar genau das gleiche ist wie die Funktion selbst, nämlich die Teilmenge des Produkts $M \times N$, die aus den Paaren (x, y) besteht, für die x bezüglich f in Relation zu y steht, also $y = f(x)$ gilt. Der Begriff des Graphen soll hier also nur noch einmal deutlich machen, dass man sich die Funktion gerade wirklich als ein derart „grafisches" Objekt vorstellt und nicht als eine „Zuordnung" von M nach N.

In der Definition 2.4 einer Abbildung $f\colon M \to N$ verlangen wir, dass jedem Element von M genau ein Element von N zugeordnet wird. Wir fordern jedoch nicht auch umgekehrt, dass jedes Element des Zielraums N das Bild von genau einem Element von M ist, oder dass es überhaupt als Bild eines Elements von M auftritt. Abbildungen, die diese Eigenschaften dennoch besitzen, haben spezielle Namen, die wir jetzt einführen wollen.

Definition 2.8 (Eigenschaften von Abbildungen)**.** Es sei $f\colon M \to N$ eine Abbildung.

(a) Ist $y \in N$ und $x \in M$ mit $f(x) = y$, so heißt x ein **Urbild** von y unter f.

(b) Hat jedes $y \in N$ ...

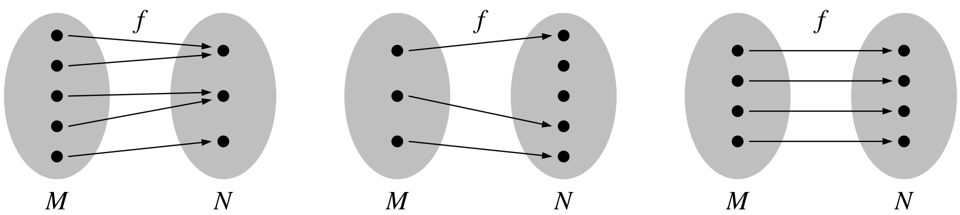
  (i) *mindestens* ein Urbild, so heißt f **surjektiv**.

  In Quantoren bedeutet dies: $\forall y \in N \; \exists x \in M\colon \; f(x) = y$.

  (ii) *höchstens* ein Urbild, so heißt f **injektiv**.

  In Quantoren bedeutet dies: $\forall x_1, x_2 \in M\colon \; f(x_1) = f(x_2) \Rightarrow x_1 = x_2$. (Also: Haben zwei Elemente des Startraums das gleiche Bild, so müssen sie bereits dasselbe Element sein.)

  (iii) *genau* ein Urbild, ist f also surjektiv und injektiv, so heißt f **bijektiv**.



Beispiel 2.9. Betrachten wir noch einmal die Funktionen aus Beispiel 2.6 (a) bzw. Bemerkung 2.7. Die Funktion f ist nicht surjektiv, da das Element 0 des Zielraums kein Urbild hat. Sie ist jedoch injektiv: Sind $x_1, x_2 \in \mathbb{R}\backslash\{0\}$ mit $f(x_1) = f(x_2)$, also $\frac{1}{x_1} = \frac{1}{x_2}$, so folgt durch Multiplikation mit $x_1 x_2$ sofort $x_1 = x_2$.

Die Funktion g dagegen ist surjektiv: Eine Zahl $y \in \mathbb{R}$ hat als Urbild $x = y$ für $y \geq 0$, und $x = y - 1$ für $y < 0$. Sie ist allerdings nicht injektiv, denn es ist $g(-1) = g(0) = 0$.

Beachte, dass diese Eigenschaften auch von der Wahl der Start- und Zielmenge abhängen: So wird z. B. f bijektiv, wenn man die Zielmenge $\mathbb{R}$ durch $\mathbb{R}\backslash\{0\}$ ersetzt.

Aufgabe 2.10. Wie viele Abbildungen gibt es zwischen den Mengen $\{1,2,3,4\}$ und $\{1,2,3,4,5,6\}$? Wie viele von ihnen sind injektiv?

Bilder und Urbilder unter Abbildungen betrachtet man oft auch von ganzen Mengen statt nur von Punkten:

Definition 2.11. Es sei $f\colon M \to N$ eine Abbildung.

(a) Für $A \subset M$ heißt die Menge

$$f(A) := \{f(x) : x \in A\} \quad \subset N$$

(also die Menge aller Bilder von Punkten in A) das **Bild** von A unter f.

(b) Ist $B \subset N$, so heißt die Menge

$$f^{-1}(B) := \{x \in M : f(x) \in B\} \quad \subset M$$

(also die Menge aller Urbilder von Punkten in B) das **Urbild** von B unter f.

Bemerkung 2.12. Die Grundidee der Notation in Definition 2.11 (a) ist: Schreiben wir als Argument einer Funktion $f\colon M \to N$ eine *Teilmenge* statt einem *Element* von M, so bedeutet dies, dass wir alle Werte $f(x)$ für $x \in M$ zusammen nehmen und diese wieder in einer Menge zusammenfassen. Diese Schreibweise verwendet man auch oft, wenn die Funktion aus einer Rechenverknüpfung besteht, wie z. B. in

$$\mathbb{N}+\tfrac{1}{2} := \left\{n+\tfrac{1}{2} : n \in \mathbb{N}\right\} = \left\{\tfrac{1}{2},\tfrac{3}{2},\tfrac{5}{2},\dots\right\} \quad \text{oder} \quad 3\mathbb{Z} := \{3n : n \in \mathbb{Z}\} = \{\dots,-6,-3,0,3,6,\dots\}.$$

Beispiel 2.13. Für die Funktion $f\colon \mathbb{R}\backslash\{0\} \to \mathbb{R}$, $x \mapsto \frac{1}{x}$ aus Beispiel 2.6 (a) ist $f(\mathbb{R}_{>0}) = \mathbb{R}_{>0}$ und $f^{-1}(\{0\}) = \emptyset$.

Beispiel 2.14. Zwischen den Konstruktionen von Bild und Urbild aus Definition 2.11 und den Mengenoperationen aus Abschnitt 1.B gibt es sehr viele Beziehungen. Um einmal exemplarisch zu sehen, wie derartige Beziehungen aussehen und bewiesen werden können, wollen wir nun zeigen, dass für jede Abbildung $f\colon M \to N$ und zwei beliebige Teilmengen $A,B \subset M$ stets

$$f(A)\backslash f(B) \subset f(A\backslash B) \tag{$*$}$$

gilt.

Zum Beweis müssen wir zeigen, dass jedes Element der linken Menge auch in der rechten Menge liegt. Es sei also $y \in f(A)\backslash f(B)$ beliebig. Insbesondere ist damit $y \in f(A)$, nach Definition 2.11 (a) also $y = f(x)$ für ein $x \in A$. Würde nun auch $x \in B$ gelten, so hätten wir wegen $y = f(x)$ auch $y \in f(B)$, im Widerspruch zu $y \in f(A)\backslash f(B)$. Also ist $x \notin B$, und damit $x \in A\backslash B$. Damit besagt $y = f(x)$ aber gerade $y \in f(A\backslash B)$. Insgesamt haben wir somit die behauptete Teilmengenbeziehung $(*)$ gezeigt.

Beachte allerdings, dass in $(*)$ im Allgemeinen keine Gleichheit gilt: Für $f\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$ mit $A = \{-1,1\}$ und $B = \{-1\}$ ist

$$f(A)\backslash f(B) = \{1\}\backslash\{1\} = \emptyset, \quad \text{aber} \quad f(A\backslash B) = f(\{1\}) = \{1\}.$$

Aufgabe 2.15. Beweise die folgenden Teilmengenbeziehungen und untersuche jeweils, ob auch die Gleichheit gilt.

(a) Für alle Mengen M,A,B gilt $M\backslash(A\cup B) \subset (M\backslash A)\cap(M\backslash B)$.

(b) Ist $f\colon M \to N$ eine Abbildung und $A \subset N$, so ist $f\left(f^{-1}(A)\right) \subset A$.

Aufgabe 2.16. Es sei $f\colon M \to N$ eine Abbildung. Finde für das Symbol $\square$ jeweils eine der Mengenbeziehungen $\subset,=,\supset$, so dass die folgenden Aussagen wahr werden, und beweise die so entstandenen Aussagen!

(a) $f(A)\cap f(B) \,\square\, f(A\cap B)$ für alle $A,B \subset M$.

(b) $f^{-1}(A) \cap f^{-1}(B) \square f^{-1}(A \cap B)$ für alle $A, B \subset N$.

Als Nächstes wollen wir nun die euch sicher bereits bekannte Verkettung, also die Hintereinanderausführung von Funktionen einführen.

Definition 2.17 (Verkettung von Funktionen)**.** Es seien $f\colon M \to N$ und $g\colon N \to R$ zwei Abbildungen (also so dass die Zielmenge von f gleich der Startmenge von g ist). Dann heißt die Abbildung

$$g \circ f\colon M \to R,\ x \mapsto g(f(x))$$

die **Verkettung** von f und g.

Bemerkung 2.18. Bei der Verkettung zweier Funktionen kommt es natürlich auf die Reihenfolge an, allein schon weil in der Situation von Definition 2.17 in der Regel der Zielraum von g ja nicht mit dem Startraum von f übereinstimmt und die „umgekehrte Verkettung" $f \circ g$ damit gar nicht definierbar wäre. Beachte dabei, dass die Notation $g \circ f$ lautet, obwohl wir zuerst f (von M nach N) und dann g (von N nach R) anwenden. Diese vielleicht etwas merkwürdig erscheinende Notation kommt einfach daher, dass die Buchstaben in der gleichen Reihenfolge stehen sollen wie bei der Abbildungsvorschrift $x \mapsto g(f(x))$.

Wir wollen nun unser erstes *Lemma* beweisen – „Lemma" ist griechisch und bedeutet eigentlich „Annahme", aber in der Mathematik wird dieser Begriff für einen *Hilfssatz* verwendet, also für ein kleines Zwischenresultat, das vielleicht für sich genommen nicht übermäßig überraschend oder interessant ist, aber das in späteren Beweisen immer wieder nützlich sein wird. In unserem momentanen Fall geht es einfach darum, dass die Verkettung von Abbildungen *assoziativ* ist (siehe auch Definition 3.1):

Lemma 2.19 (Assoziativität der Verkettung)**.** *Sind $f\colon M \to N$, $g\colon N \to R$ und $h\colon R \to S$ drei Abbildungen, so gilt $h \circ (g \circ f) = (h \circ g) \circ f$. (Man schreibt für diese Abbildung daher oft auch einfach $h \circ g \circ f$.)*

Beweis. Nach Definition 2.4 können wir die Gleichheit zweier Funktionen zeigen, indem wir für jedes Element der Startmenge nachweisen, dass sein Bild unter beiden Funktionen übereinstimmt. Dies rechnen wir nun einfach durch wiederholtes Einsetzen von Definition 2.17 nach: Es gilt

$$(h \circ (g \circ f))(x) = h((g \circ f)(x)) = h(g(f(x)))$$

und

$$((h \circ g) \circ f)(x) = (h \circ g)(f(x)) = h(g(f(x))).$$

Da diese beiden Ausdrücke übereinstimmen, ist das Lemma bewiesen. □

Schließlich wollen wir nun noch sehen, wie sich die Begriffe der Surjektivität, Injektivität und Bijektivität aus Definition 2.8 (b) äquivalent mit Hilfe von Umkehrfunktionen ausdrücken lassen.

Lemma 2.20 (Umkehrfunktionen)**.** *Es sei $f\colon M \to N$ eine Abbildung zwischen nicht-leeren Mengen. Dann gilt:*

(a) *f ist surjektiv $\Leftrightarrow$ es gibt eine Abbildung $g\colon N \to M$ mit $f \circ g = \mathrm{id}_N$.*

(b) *f ist injektiv $\Leftrightarrow$ es gibt eine Abbildung $g\colon N \to M$ mit $g \circ f = \mathrm{id}_M$.*

(c) *f ist bijektiv $\Leftrightarrow$ es gibt eine Abbildung $g\colon N \to M$ mit $f \circ g = \mathrm{id}_N$ und $g \circ f = \mathrm{id}_M$.*

In diesem Fall ist diese Abbildung g eindeutig bestimmt. Wir nennen sie die ***Umkehrfunktion*** *bzw.* ***Umkehrabbildung*** *von f und bezeichnen sie mit f^{-1}.*

Beweis. Wir beweisen die Richtungen „$\Rightarrow$" und „$\Leftarrow$" der drei Teile separat.

(a) „$\Rightarrow$": Es sei f surjektiv. Dann können wir

$$g\colon N \to M,\ y \mapsto \text{ein Urbild von } y \text{ unter } f$$

setzen (wobei wir, falls mehrere solche Urbilder existieren, jeweils eines auswählen). Für alle $y \in N$ gilt dann $f(g(y)) = y$, da $g(y)$ ja ein Urbild von y unter f ist. Also ist $f \circ g = \mathrm{id}_N$.

„$\Leftarrow$“: Es gebe ein $g\colon N \to M$ mit $f \circ g = \mathrm{id}_N$. Ist dann $y \in N$ beliebig, so können wir $x = g(y)$ setzen, und es gilt $f(x) = f(g(y)) = \mathrm{id}_N(y) = y$. Also hat jedes solche y ein Urbild unter f, d. h. f ist surjektiv.

(b) „$\Rightarrow$“: Es sei f injektiv. Dann setzen wir

$$g\colon N \to M,\ y \mapsto \begin{cases} \text{das eindeutig bestimmte Urbild von } y \text{ unter } f, \text{ falls eines existiert,} \\ \text{ein beliebiges Element von } M \text{ sonst} \end{cases}$$

(wobei wir in der zweiten Zeile die Voraussetzung $M \neq \emptyset$ brauchen). Für alle $x \in M$ ist dann $g(f(x))$ nach Konstruktion das eindeutige Urbild von $f(x)$ unter f, also x. Damit folgt $g \circ f = \mathrm{id}_M$.

„$\Leftarrow$“: Es gebe ein $g\colon N \to M$ mit $g \circ f = \mathrm{id}_M$. Wir müssen zeigen, dass f injektiv ist, also die Bedingung aus Definition 2.8 (b)(ii) überprüfen. Es seien dazu $x_1, x_2 \in M$ mit $f(x_1) = f(x_2)$. Anwenden von g liefert $g(f(x_1)) = g(f(x_2))$, wegen $g \circ f = \mathrm{id}_M$ also $x_1 = x_2$. Damit ist f injektiv.

(c) „$\Rightarrow$“: Ist f bijektiv, so können wir

$$g\colon N \to M,\ y \mapsto \text{das eindeutig bestimmte Urbild von } y \text{ unter } f$$

setzen. Da diese Definition zu denen aus (a) und (b) „$\Rightarrow$“ passt, zeigen die obigen Rechnungen also, dass f dann surjektiv und injektiv und damit bijektiv ist.

„$\Leftarrow$“: Dies folgt sofort aus Teil „$\Leftarrow$“ von (a) und (b).

Eindeutigkeit von g: Es seien $g_1, g_2\colon N \to M$ zwei Umkehrfunktionen von f, insbesondere ist also $f \circ g_2 = \mathrm{id}_N$ und $g_1 \circ f = \mathrm{id}_M$. Dann gilt

$$g_1 = g_1 \circ \mathrm{id}_N = g_1 \circ (f \circ g_2) \overset{2.19}{=} (g_1 \circ f) \circ g_2 = \mathrm{id}_M \circ g_2 = g_2. \qquad \square$$

Bemerkung 2.21 (Urbilder und Umkehrabbildungen)**.** Beachte, dass wir das Urbild einer Menge unter einer Abbildung $f\colon M \to N$ in Definition 2.11 (b) mit dem gleichen Symbol f^{-1} bezeichnet haben wie (im Fall einer bijektiven Abbildung) die Umkehrabbildung aus Lemma 2.20 (c). Das ist vielleicht etwas unglücklich gewählt, aber in der Literatur so fest verankert, dass wir hier nicht davon abweichen wollen. Bei genauem Hinschauen kann man aber auch immer feststellen, was gemeint ist:

> Ist $f\colon M \to N$ eine Abbildung, so bezeichnet ...
>
> ... $f^{-1}(B)$ für eine *Menge* $B \subset N$ das *Urbild* von B wie in Definition 2.11 (b); es existiert für jede Abbildung f.
>
> ... $f^{-1}(y)$ für ein *Element* $y \in B$ den *Wert der Umkehrabbildung* bei y wie in Lemma 2.20 (c); er existiert nur für bijektives f.

Letztlich hängen diese beiden Notationen aber auch eng miteinander zusammen: Ist f bijektiv und ist $x \in M$ mit $f(x) = y$, so ist $f^{-1}(y) = x$ (mit f^{-1} im Sinne der Umkehrabbildung) und $f^{-1}(\{y\}) = \{x\}$ (mit f^{-1} im Sinne des Urbildes).

Aufgabe 2.22.

(a) Untersuche die Abbildung $f\colon \mathbb{Z} \to \mathbb{Z},\ x \mapsto 3x+2$ auf Injektivität und Surjektivität.

(b) Untersuche die Abbildung $f\colon \mathbb{R}^2 \to \mathbb{R}^2,\ (x,y) \mapsto (xy, x+1)$ auf Injektivität und Surjektivität.

(c) Man zeige: Sind $f\colon M \to N$ und $g\colon N \to R$ surjektiv, so ist auch $g \circ f\colon M \to R$ surjektiv.

Aufgabe 2.23. Es seien $f\colon M \to N$ und $g\colon N \to R$ bijektiv. Zeige, dass dann auch $f^{-1}\colon N \to M$ und $g \circ f\colon M \to R$ bijektiv sind.

Aufgabe 2.24. Man beweise oder widerlege:

(a) Sind $f\colon M \to N$ und $g\colon N \to R$ zwei Abbildungen und ist $g \circ f$ injektiv, so ist auch f injektiv.

(b) Sind $f\colon M \to N$ und $g\colon N \to R$ zwei Abbildungen und ist $g \circ f$ injektiv, so ist auch g injektiv.

Aufgabe 2.25. Es sei $f: M \to N$ eine Abbildung zwischen zwei Mengen. Man beweise:

$$f \text{ ist surjektiv} \quad \Leftrightarrow \quad \text{für alle } A,B \subset N \text{ mit } f^{-1}(A) = f^{-1}(B) \text{ gilt } A = B.$$

2.B Äquivalenzrelationen

Am Anfang dieses Kapitels haben wir allgemeine Relationen eingeführt, als einzigen Spezialfall davon aber bisher nur die Funktionen ausführlicher betrachtet. Wir wollen daher nun noch einen ganz anderen wichtigen Typ von Relationen studieren, die sogenannten Äquivalenzrelationen.

Angenommen, wir möchten eine Menge M untersuchen, die uns zunächst einmal zu groß oder zu kompliziert erscheint. Es gibt dann zwei prinzipiell verschiedene Möglichkeiten, wie man daraus eine kleinere bzw. einfachere Menge machen kann:

- Wir können uns auf eine Teilmenge von M beschränken – dann schließen wir allerdings manche Elemente von M von unserer Untersuchung aus.
- Wir können zwar alle Elemente von M betrachten, aber manche von ihnen miteinander identifizieren bzw. zu Klassen zusammenfassen – d. h. sie als gleich bzw. „äquivalent“ ansehen, wenn sie für das zu untersuchende Problem ähnliche Eigenschaften haben.

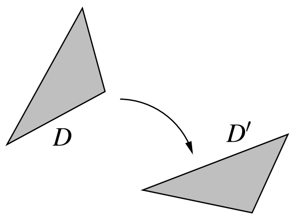
Diese zweite Idee der Identifizierung ähnlicher Elemente führt zum Begriff der Äquivalenzrelationen. Sie klingt vielleicht zunächst etwas abstrakt, ist euch aber sicher schon an vielen Stellen begegnet. Hier sind zwei einfache Beispiele dafür.

Beispiel 2.26.

(a) Eine analoge Uhr vereinfacht die recht große Menge aller (vergangenen und zukünftigen) Zeitpunkte, die wir uns als Zeitachse $M = \mathbb{R}$ vorstellen können, indem sie nach jeweils 12 Stunden wieder dasselbe anzeigt. Sie identifiziert also zwei Zeitpunkte $x, y \in \mathbb{R}$ (gemessen in Stunden) miteinander, wenn $x - y$ ein Vielfaches von 12 ist. Dadurch „verkleinert“ sie die ursprüngliche Zeitachse auf ein gut überschaubares Intervall von 12 Stunden – und wir alle wissen, dass uns ein Blick auf die Uhr in vielen Fällen ausreicht, wenn wir den aktuellen Zeitpunkt wissen wollen, auch wenn uns das nichts über das Datum oder die Tageszeit (vormittags oder nachmittags) sagt.

(b) Als „mathematischeres“ Beispiel können wir die Menge M aller Dreiecke in der Ebene $\mathbb{R}^2$ betrachten.

Bekanntlich heißen zwei solche Dreiecke $D, D' \in M$ zueinander *kongruent*, wenn sie wie im Bild rechts durch eine Drehung und / oder Verschiebung auseinander hervorgehen – wir schreiben dies im Folgenden als $D \sim D'$. Zueinander kongruente Dreiecke werden oft miteinander identifiziert, nämlich immer dann, wenn es uns nur auf die Form bzw. Größe der Dreiecke, aber nicht auf ihre Lage in der Ebene ankommt.



Wenn wir z. B. sagen, dass die drei Seitenlängen ein Dreieck eindeutig bestimmen, dann meinen wir damit in Wirklichkeit, dass sie das Dreieck *bis auf Kongruenz* eindeutig bestimmen, also nur die Form und Größe festlegen, aber nicht die Lage des Dreiecks in $\mathbb{R}^2$. Formal kann man dies so ausdrücken: zu einem Dreieck D nennt man

$$\overline{D} := \{D' \in M : D' \sim D\},$$

also die Menge aller zu D kongruenten Dreiecke, die *Kongruenzklasse* von D. Die Menge aller dieser Kongruenzklassen bezeichnen wir mit

$$M/\sim \; := \; \{\overline{D} : D \in M\}.$$

Man kann dann z. B. sagen, dass die Seitenlängen eines Dreiecks ein eindeutiges Element in $M/\sim$ bestimmen, also eine eindeutige Kongruenzklasse von Dreiecken festlegen – nicht aber ein eindeutiges Element von M.

Mit der Idee dieser Beispiele im Kopf wollen wir nun den Begriff der Äquivalenzrelation exakt definieren.

Definition 2.27 (Äquivalenzrelationen)**.** Es sei $\sim$ wie in Definition 2.1 eine Relation auf einer Menge M. Wie in Bemerkung 2.2 schreiben wir $x \sim y$, wenn x und y bezüglich $\sim$ in Relation stehen.

Man nennt $\sim$ eine **Äquivalenzrelation** auf M, wenn die folgenden Eigenschaften gelten:

(a) **Reflexivität**: Für alle $x \in M$ gilt $x \sim x$.

(b) **Symmetrie**: Sind $x, y \in M$ mit $x \sim y$, so gilt auch $y \sim x$.

(c) **Transitivität**: Sind $x, y, z \in M$ mit $x \sim y$ und $y \sim z$, so gilt auch $x \sim z$.

In diesem Fall sagt man statt $x \sim y$ auch, dass x (bezüglich dieser Relation) zu y **äquivalent** ist. Zu $x \in M$ heißt dann die Menge

$$\overline{x} := \{y \in M : y \sim x\}$$

aller Elemente, die zu x äquivalent sind, die **Äquivalenzklasse** bzw. einfach **Klasse** von x; jedes Element dieser Menge nennt man einen **Repräsentanten** dieser Klasse. Die Menge aller Äquivalenzklassen schreiben wir als

$$M/\sim \;:=\; \{\overline{x} : x \in M\}.$$

Beispiel 2.28.

(a) Das Beispiel 2.26 (a) einer analogen Uhr lässt sich mathematisch exakt wie folgt definieren: Auf $M = \mathbb{R}$ betrachten wir die Relation

$$x \sim y \;:\Leftrightarrow\; \text{es gibt ein } k \in \mathbb{Z} \text{ mit } x - y = 12k. \qquad (*)$$

Dies ist in der Tat eine Äquivalenzrelation, denn für alle $x, y, z \in \mathbb{R}$ gilt:

- Reflexivität: Es gilt $x - x = 12 \cdot 0 = 12k$ mit $k = 0 \in \mathbb{Z}$, also $x \sim x$.
- Symmetrie: Es gelte $x \sim y$, also $x - y = 12k$ für ein $k \in \mathbb{Z}$. Durch Multiplikation mit -1 folgt dann auch $y - x = 12 \cdot (-k) = 12k'$ mit $k' := -k \in \mathbb{Z}$, und damit $y \sim x$.
- Transitivität: Es gelte $x \sim y$ und $y \sim z$, nach Definition der Relation also $x - y = 12k$ und $y - z = 12k'$ für gewisse $k, k' \in \mathbb{Z}$ (beachte, dass der Wert von k in $(*)$ von x und y abhängt und wir daher für die Differenz $y - z$ eine neue Variable k' brauchen). Durch Addition dieser beiden Gleichungen erhalten wir $x - z = 12(k + k') = 12k''$ mit $k'' := k + k' \in \mathbb{Z}$, und damit $x \sim z$.

Die Äquivalenzklasse z. B. von $2 \in \mathbb{R}$ ist

$$\overline{2} = \{x \in \mathbb{R} : x \sim 2\} = \{x \in \mathbb{R} : \text{es gibt ein } k \in \mathbb{Z} \text{ mit } x - 2 = 12k\} = \{2 + 12k : k \in \mathbb{Z}\},$$

also die Menge aller Zeitpunkte, zu denen die Uhr auf 2 steht. Jeder beliebige Zeitpunkt $x \in \mathbb{R}$, zu dem die Uhr auf 2 steht, ist ein Repräsentant dieser Klasse, und die Menge $M/\sim$ entspricht allen möglichen Ständen der Uhr.

(b) Die Kongruenz von Dreiecken aus Beispiel 2.26 (b) ist ebenfalls eine Äquivalenzrelation (es ist offensichtlich, dass sie die Eigenschaften aus Definition 2.27 erfüllt). Die Äquivalenzklassen sind in diesem Fall genau die Kongruenzklassen.

(c) Die „Kleiner-Relation“ auf $\mathbb{R}$ aus Beispiel 2.3 (b), also die Relation, für die für $x, y \in \mathbb{R}$ genau dann $x \sim y$ gilt, wenn $x < y$ ist, ist keine Äquivalenzrelation, da sie weder reflexiv noch symmetrisch ist.

Beachte, dass bei unseren Äquivalenzrelationen aus Beispiel 2.28 (a) und (b) jedes Element von M in genau einer Äquivalenzklasse liegt: Zu jedem Zeitpunkt hat eine analoge Uhr genau einen Stand, und jedes Dreieck in der Ebene liegt in genau einer Kongruenzklasse. Dies beschreibt genau unsere Idee, dass wir die Elemente von M auf eine bestimmte Art zu Klassen zusammenfassen wollen. Allgemein sind die Axiome einer Äquivalenzrelation aus Definition 2.27 anschaulich genau diejenigen, die man braucht, damit die Relation sinnvoll eine solche Identifizierung von Elementen zu Klassen beschreiben kann. Dies zeigt auch noch einmal der folgende zentrale Satz über Äquivalenzrelationen.

Satz 2.29 (Eigenschaften von Äquivalenzrelationen)**.** *Es sei $\sim$ eine Äquivalenzrelation auf einer Menge M.*

(a) *Für $x, y \in M$ gilt $x \sim y$ genau dann, wenn $\overline{x} = \overline{y}$. (Zwei Elemente sind also genau dann äquivalent zueinander, wenn sie die gleiche Äquivalenzklasse bestimmen.)*

(b) *Jedes Element $x \in M$ liegt in genau einer Äquivalenzklasse (nämlich in $\overline{x}$). Insbesondere ist M also die disjunkte Vereinigung aller Äquivalenzklassen. Man sagt auch, dass die Äquivalenzklassen eine* Partition *von M bilden.*

Beweis.

(a) Es seien $x, y \in M$.

„$\Rightarrow$": Es gelte $x \sim y$. Ist dann $z \in M$ mit $z \in \overline{x}$, also $z \sim x$, so ist nach der Transitivität wegen $x \sim y$ auch $z \sim y$, also $z \in \overline{y}$. Damit gilt $\overline{x} \subset \overline{y}$. Da mit $x \sim y$ wegen der Symmetrie aber auch $y \sim x$ gilt, folgt analog auch umgekehrt $\overline{y} \subset \overline{x}$, und somit insgesamt $\overline{x} = \overline{y}$.

„$\Leftarrow$": Es sei nun $\overline{x} = \overline{y}$. Wegen der Reflexivität ist $x \sim x$, also $x \in \overline{x} = \overline{y}$, und damit $x \sim y$.

(b) Wegen der Reflexivität liegt natürlich jedes $x \in M$ in seiner eigenen Äquivalenzklasse $\overline{x}$. Ist nun auch $x \in \overline{y}$ für ein $y \in M$, also $x \sim y$, so gilt nach (a) bereits $\overline{y} = \overline{x}$. Also liegt x in genau einer Äquivalenzklasse von $\sim$, nämlich in $\overline{x}$. $\square$

Aufgabe 2.30. Es sei $M = \{n \in \mathbb{Z} : |n| \le 100\} = \{-100, -99, \dots, 0, \dots, 99, 100\}$. Welche der folgenden Relationen sind Äquivalenzrelationen auf M? Gib im Fall einer Äquivalenzrelation außerdem die Äquivalenzklasse $\overline{-34}$ explizit an.

(a) $x \sim y \;:\Leftrightarrow\;$ es gibt ein $n \in \mathbb{Z}$ mit $x = 2^n y$.

(b) $x \sim y \;:\Leftrightarrow\; xy \ge 0$.

Aufgabe 2.31. Welche der folgenden Relationen sind Äquivalenzrelationen auf $\mathbb{R}^2$? Im Fall einer Äquivalenzrelation berechne und skizziere man außerdem die Äquivalenzklassen von $(0,1) \in \mathbb{R}^2$ und $(1,1) \in \mathbb{R}^2$.

(a) $(x,y) \sim (x',y') \;:\Leftrightarrow\;$ es gibt ein $a \in \mathbb{R}\backslash\{0\}$ mit $x = a^2 x'$ und $y = ay'$;

(b) $(x,y) \sim (x',y') \;:\Leftrightarrow\;$ es gibt ein $a \in \mathbb{R}\backslash\{0\}$ mit $x = ay'$ und $y = ax'$.

3. Erste Eigenschaften der reellen Zahlen

In Notation 1.15 haben wir bereits die reellen Zahlen $\mathbb{R}$ als „Menge der Punkte auf einer Geraden“ eingeführt. Man kann aber natürlich noch viel mehr Dinge mit den reellen Zahlen tun als sie als eine einfache Punktmenge zu betrachten: Man kann sie addieren, multiplizieren, die Größe von zwei Zahlen miteinander vergleichen, und noch einiges mehr. Wir wollen die Eigenschaften der reellen Zahlen in diesem und dem nächsten Kapitel exakt formalisieren, damit wir danach genau wissen, welche Eigenschaften von $\mathbb{R}$ wir in dieser Vorlesung axiomatisch voraussetzen. In der Tat werden diese Eigenschaften letztlich sogar ausreichen, um die reellen Zahlen eindeutig zu charakterisieren. Wir beginnen in diesem Kapitel aber zunächst einmal nur mit den „Grundrechenarten“, also mit der Addition und der Multiplikation sowie ihren Umkehrungen, der Subtraktion und Division.

3.A Gruppen und Körper

Die Eigenschaften von Verknüpfungen wie der Addition oder Multiplikation reeller Zahlen werden mathematisch durch die Begriffe einer Gruppe bzw. eines Körpers beschrieben, die wir jetzt einführen wollen.

Definition 3.1 (Gruppen)**.** Eine **Gruppe** ist eine Menge G zusammen mit einer „Verknüpfung“, d. h. einer Abbildung

$$*\colon G\times G\to G,\ (x,y)\mapsto x*y,$$

so dass die folgenden Eigenschaften (auch *Gruppenaxiome* genannt) gelten:

(a) (**Assoziativität**) Für alle $x,y,z\in G$ gilt $(x*y)*z=x*(y*z)$. Man schreibt diesen Ausdruck dann in der Regel auch einfach als $x*y*z$, weil die Reihenfolge der Klammerung ja egal ist.

(b) (Existenz eines neutralen Elements) Es gibt ein $e\in G$, für das $e*x=x*e=x$ für alle $x\in G$ gilt. Man nennt ein solches e ein **neutrales Element**, und verlangt davon zusätzlich:

(c) (Existenz von inversen Elementen) Für alle $x\in G$ gibt es ein $x'\in G$ mit $x'*x=x*x'=e$. Man nennt x' dann ein **inverses Element** zu x.

Wir bezeichnen eine solche Gruppe mit $(G,*)$. Wenn aus dem Zusammenhang klar ist, welche Verknüpfung gemeint ist, schreiben wir oft auch einfach nur G für die Gruppe.

Gilt zusätzlich zu den obigen Eigenschaften noch

(d) (**Kommutativität**) $x*y=y*x$ für alle $x,y\in G$,

so heißt $(G,*)$ eine **kommutative** oder **abelsche Gruppe**.

Bemerkung 3.2. Manchmal wird in der Definition einer Gruppe in Teil (b) lediglich $e*x=x$ und in Teil (c) lediglich $x'*x=e$ gefordert (man spricht dann auch von einem **linksneutralen** bzw. **linksinversen** Element). Man kann jedoch unter Verwendung der übrigen Gruppenaxiome zeigen, dass in diesem Fall automatisch auch $x*e=x$ und $x*x'=e$ gelten muss, also dass linksneutrale Elemente immer neutral und linksinverse Element bereits immer inverse Elemente sind [G, Satz 1.7]. Die beiden Varianten der Definition einer Gruppe stimmen also letztlich überein.

Beispiel 3.3.

(a) $(\mathbb{R},+)$ ist eine abelsche Gruppe, denn die Addition ist (wie wir axiomatisch voraussetzen werden) eine Verknüpfung auf $\mathbb{R}$ mit den Eigenschaften:

- $(x+y)+z=x+(y+z)$ für alle $x,y,z\in\mathbb{R}$;
- $0\in\mathbb{R}$ ist ein neutrales Element, denn $0+x=x+0=x$ für alle $x\in\mathbb{R}$;
- zu jedem $x\in\mathbb{R}$ ist $-x\in\mathbb{R}$ ein inverses Element, denn $(-x)+x=x+(-x)=0$;

- $x+y=y+x$ für alle $x,y\in\mathbb{R}$.

Auf die gleiche Art sind auch $(\mathbb{Q},+)$ und $(\mathbb{Z},+)$ abelsche Gruppen, jedoch nicht $(\mathbb{N},+)$: Hier existiert zwar noch ein neutrales Element 0, aber die Zahl $1\in\mathbb{N}$ hat kein Inverses mehr, denn es gibt kein $x\in\mathbb{N}$ mit $x+1=0$.

(b) $(\mathbb{R},\cdot)$ ist keine Gruppe: Die Multiplikation ist zwar assoziativ und kommutativ und hat das neutrale Element 1, aber die Zahl 0 hat kein Inverses – denn dies müsste ja eine Zahl $x\in\mathbb{R}$ sein mit $x\cdot 0=1$.

Nimmt man jedoch die 0 aus $\mathbb{R}$ heraus, so erhält man mit $(\mathbb{R}\backslash\{0\},\cdot)$ wieder eine abelsche Gruppe, bei der das neutrale Element 1 und das zu einem x inverse Element $\frac{1}{x}$ ist. Genauso funktioniert dies für $(\mathbb{Q}\backslash\{0\},\cdot)$, aber z. B. nicht für $(\mathbb{Z}\backslash\{0\},\cdot)$: Hier gibt es zwar noch ein neutrales Element 1, aber die Zahl $2\in\mathbb{Z}\backslash\{0\}$ hat kein Inverses mehr, denn es gibt kein $x\in\mathbb{Z}\backslash\{0\}$ mit $2\cdot x=1$.

(c) Hier ist noch ein Beispiel von einem ganz anderen Typ: Es sei M eine beliebige Menge und $G=\{f\colon M\to M \text{ bijektiv}\}$ die Menge aller bijektiven Abbildungen von M nach M. Da die Verkettung bijektiver Abbildungen nach Aufgabe 2.23 wieder bijektiv ist, definiert sie eine Verknüpfung auf G. In der Tat wird G damit zu einer Gruppe, denn die Verkettung ist assoziativ nach Lemma 2.19, die Identität id_M ist ein neutrales Element, und zu einem $f\in G$ ist die Umkehrabbildung f^{-1} aus Lemma 2.20 (c) ein inverses Element: Sie ist nach Aufgabe 2.23 selbst wieder bijektiv (also in G) und erfüllt $f^{-1}\circ f=f\circ f^{-1}=\mathrm{id}_M$ nach Lemma 2.20 (c). Im Allgemeinen ist diese Gruppe jedoch nicht kommutativ.

Wir wollen nun ein paar einfache Eigenschaften von Gruppen beweisen, u. a. dass die in Definition 3.1 geforderten neutralen und inversen Elemente eindeutig sind und wir daher in Zukunft auch von *dem* neutralen und *dem* zu einem gegebenen Element inversen Element sprechen können.

Lemma 3.4 (Eigenschaften von Gruppen)**.** *Es seien $(G,*)$ eine Gruppe und $x,y\in G$.*

(a) *Es gibt genau ein neutrales Element (wie in Definition 3.1 (b)).*

(b) *Es gibt genau ein inverses Element zu x (wie in Definition 3.1 (c)).*

(c) *Sind x' und y' die inversen Elemente zu x bzw. y, so ist $y'*x'$ das inverse Element zu $x*y$.*

(d) *Ist x' das inverse Element zu x, so ist x das inverse Element zu x' („das Inverse des Inversen ist wieder das Ausgangselement“).*

Beweis.

(a) Sind e und $\tilde{e}$ neutrale Elemente, so folgt

$$\begin{aligned} e &= \tilde{e}*e && \text{(denn } \tilde{e} \text{ ist ein neutrales Element)}\\ &= \tilde{e} && \text{(denn } e \text{ ist ein neutrales Element).} \end{aligned}$$

(b) Sind x' und $\tilde{x}'$ inverse Elemente zu x, so gilt

$$\begin{aligned} x' &= e*x' && (e \text{ neutrales Element})\\ &= (\tilde{x}'*x)*x' && (\tilde{x}' \text{ ist ein inverses Element zu } x)\\ &= \tilde{x}'*(x*x') && \text{(Assoziativität)}\\ &= \tilde{x}'*e && (x' \text{ ist ein inverses Element zu } x)\\ &= \tilde{x}' && (e \text{ neutrales Element}). \end{aligned}$$

(c) Es gilt

$$(y'*x')*(x*y)=y'*(x'*x)*y=y'*e*y=y'*y=e$$

und analog auch $(x*y)*(y'*x')=e$. Damit ist $y'*x'$ das inverse Element zu $x*y$.

(d) Die Gleichung $x'*x=x*x'=e$ besagt direkt, dass x das inverse Element zu x' ist. □

Wie wir in Beispiel 3.3 (a) und (b) gesehen haben, erlauben die reellen Zahlen *zwei* grundlegende Gruppenstrukturen: die Addition und (nach Herausnahme der 0) die Multiplikation. Diese beiden Strukturen sind jedoch nicht unabhängig voneinander, da sie durch das Distributivgesetz $(x+y)\cdot z = xz+yz$ für alle $x,y,z \in \mathbb{R}$ miteinander verbunden sind. Eine derartige Kombination zweier Gruppenstrukturen bezeichnet man als einen Körper.

Definition 3.5 (Körper)**.** Ein **Körper** ist eine Menge K zusammen mit zwei Verknüpfungen

$$+ \colon K \times K \to K \quad \text{(genannt } \textit{Addition}\text{)} \qquad \text{und} \qquad \cdot \colon K \times K \to K \quad \text{(genannt } \textit{Multiplikation}\text{)},$$

so dass die folgenden Eigenschaften (auch *Körperaxiome* genannt) gelten:

(a) $(K,+)$ ist eine abelsche Gruppe. Wir bezeichnen ihr neutrales Element mit 0 und das zu einem $x \in K$ inverse Element mit $-x$.

(b) $(K\backslash\{0\},\cdot)$ ist ebenfalls eine abelsche Gruppe. Wir bezeichnen ihr neutrales Element mit 1 und das zu einem $x \in K\backslash\{0\}$ inverse Element mit x^{-1}.

(c) (Distributivität) Für alle $x,y,z \in K$ gilt $(x+y)\cdot z = (x\cdot z)+(y\cdot z)$.

Mit dieser Definition wollen wir nun also axiomatisch voraussetzen:

$$\boxed{\mathbb{R} \text{ ist ein Körper.}}$$

Um Verwirrungen zu vermeiden, werden wir die beiden Verknüpfungen in einem Körper immer mit den Symbolen „+“ und „·“ bezeichnen. Ebenso werden wir (wie ihr es natürlich gewohnt seid) vereinbaren, dass man den Punkt bei der Multiplikation auch weglassen darf und bei ungeklammerten Ausdrücken zuerst die Multiplikationen und dann die Additionen ausgeführt werden, so dass man also z. B. die Distributivität aus Definition 3.5 (c) auch als $(x+y)z = xz+yz$ schreiben kann.

Es ist jedoch wichtig zu verstehen, dass wir ab jetzt *nicht* mehr voraussetzen werden, dass Addition und Multiplikation in einem Körper wie z. B. $\mathbb{R}$ genau die Verknüpfungen sind, „an die man als Erstes denken würde“ – was auch immer das heißen mag. Stattdessen sind es einfach irgendwelche zwei Verknüpfungen, die die Eigenschaften aus Definition 3.5 haben. Unsere zukünftigen Beweise über Körper wie z. B. $\mathbb{R}$ müssen wir also ausschließlich auf diesen Eigenschaften aufbauen.

Dieser axiomatische Zugang hat zwei Vorteile:

- Zum einen wissen wir dadurch genau, welche Eigenschaften der Grundrechenarten auf den reellen Zahlen wir eigentlich voraussetzen. Es sollte schließlich klar sein, dass wir eine *exakte* Mathematik nicht auf einer *anschaulichen* Vorstellung von $\mathbb{R}$ aufbauen können. Solltet ihr euch also z. B. später einmal dafür interessieren, wie man die Existenz der reellen Zahlen beweisen kann, so wüsstet ihr dann genau, was eigentlich zu beweisen ist: nämlich die Existenz einer Menge mit genau den Eigenschaften, die wir jetzt axiomatisch voraussetzen.
- Zum anderen werdet ihr im Laufe eures Studiums noch viele weitere Körper kennenlernen, z. B. in Kapitel **??** den sehr wichtigen Körper der komplexen Zahlen. Alle Resultate, die nur auf den Körperaxiomen aufbauen, übertragen sich dann also sofort auf diese neuen Fälle, ohne dass man sich darüber noch einmal neu Gedanken machen muss.

Beispiel 3.6.

(a) Neben $\mathbb{R}$ ist auch $\mathbb{Q}$ (mit den gleichen Verknüpfungen wie auf $\mathbb{R}$) ein Körper. Die ganzen Zahlen $\mathbb{Z}$ bilden mit diesen Verknüpfungen jedoch keinen Körper, da $(\mathbb{Z}\backslash\{0\},\cdot)$ nach Beispiel 3.3 (b) keine Gruppe ist. Ebenso ist $\mathbb{N}$ mit diesen Verknüpfungen kein Körper, da hier nach Beispiel 3.3 (a) bereits die Addition keine Gruppenstruktur liefert.

(b) Hier ist ein Beispiel für einen Körper, der sich ganz anders verhält als $\mathbb{R}$ und $\mathbb{Q}$. Wir definieren auf der Menge $K = \{g,u\}$ zwei Verknüpfungen durch die folgenden Tabellen.

$$\begin{array}{c|cc} + & g & u \\ \hline g & g & u \\ u & u & g \end{array} \qquad \text{und} \qquad \begin{array}{c|cc} \cdot & g & u \\ \hline g & g & g \\ u & g & u \end{array}$$

Die Idee dieser Verknüpfungen ist, dass g für gerade und u für ungerade ganze Zahlen steht. So haben wir in der Tabelle z. B. $g+u$ als u definiert, weil die Addition einer geraden und einer ungeraden Zahl eine ungerade Zahl ergibt.

Man kann zeigen, dass K mit diesen beiden Verknüpfungen einen Körper bildet. Er wird in der Literatur mit $\mathbb{Z}_2$ bezeichnet, da seine Elemente die Reste ganzer Zahlen bei Division durch 2 beschreiben. Um zu beweisen, dass $\mathbb{Z}_2$ ein Körper ist, könnte man z. B. einfach die geforderten Eigenschaften für alle Elemente – es gibt ja nur zwei – explizit nachprüfen. In der Vorlesung „Algebraische Strukturen" zeigt man allerdings, dass man die Körperaxiome hier auch viel eleganter direkt aus den Eigenschaften von $\mathbb{Z}$ folgern kann [G, Satz 7.10]. Wir wollen uns hier damit begnügen, die neutralen und inversen Elemente anzugeben:

- Das additive neutrale Element ist g, wie man leicht aus der Tabelle abliest. Im Sinne der Notationen von Definition 3.5 ist also $0=g$. Wegen $g+g=u+u=g=0$ sind die additiven inversen Elemente $-g=g$ und $-u=u$. Dies stimmt natürlich auch mit der Interpretation als gerade und ungerade Zahlen überein, da das Negative von einer geraden bzw. ungeraden Zahl ebenfalls wieder gerade bzw. ungerade ist.
- Das multiplikative neutrale Element in $\mathbb{Z}_2\backslash\{0\}$ ist u – in der Tat ist es ja auch das einzige Element in $\mathbb{Z}_2\backslash\{0\}$. Gemäß der Notation von Definition 3.5 ist also $1=u$.

Beachte, dass in diesem Körper $\mathbb{Z}_2$ die Gleichung $1+1=u+u=g=0$ gilt. Die Körperaxiome lassen es also zu, dass man bei fortgesetzter Addition der 1 irgendwann wieder zur 0 zurück kommt. Wir werden in dieser Vorlesung nicht viel mit dem Körper $\mathbb{Z}_2$ zu tun haben – wir haben ihn hier nur als Beispiel dafür angegeben, dass die Körperaxiome noch weit davon entfernt sind, die rationalen oder reellen Zahlen eindeutig zu charakterisieren.

Anschaulich kann man die Körperaxiome so interpretieren, dass ein Körper eine Menge ist, auf der „die vier Grundrechenarten existieren und die erwarteten Eigenschaften haben". Wir wollen nun noch ein paar weitere dieser erwarteten Eigenschaften zeigen, die bereits aus den Körperaxiomen folgen und die wir dann beim Rechnen z. B. in $\mathbb{R}$ natürlich ständig benutzen werden.

Bemerkung 3.7. Es seien K ein Körper und $x,y\in K$.

(a) Wenden wir Lemma 3.4 (c) und (d) auf die (kommutative) Addition und Multiplikation an, so sehen wir sofort, dass

$$-(x+y)=(-x)+(-y) \quad \text{und} \quad -(-x)=x$$

sowie für $x,y\neq 0$

$$(xy)^{-1}=x^{-1}\cdot y^{-1} \quad \text{und} \quad (x^{-1})^{-1}=x.$$

(b) Etwas versteckt in Definition 3.5 steht in Teil (b) u. a. die Aussage, dass die Multiplikation überhaupt eine Verknüpfung auf $K\backslash\{0\}$ ist, also dass für $x,y\in K\backslash\{0\}$ auch $xy\in K\backslash\{0\}$ gilt. Äquivalent dazu bedeutet das:

$$\text{Ist } xy=0, \text{ so gilt } x=0 \text{ oder } y=0.$$

Lemma 3.8 (Eigenschaften von Körpern)**.** *In jedem Körper K gilt für alle $x,y\in K$:*

(a) $0\cdot x=0$.

(b) $x\cdot(-y)=-(xy)$.

(c) *Für $x\neq 0$ ist $-(x^{-1})=(-x)^{-1}$.*

Beweis.

(a) Es gilt

$$\begin{aligned} 0\cdot x &= (0+0)\cdot x && \text{(0 ist additives neutrales Element)} \\ &= 0\cdot x+0\cdot x, && \text{(Distributivität)} \end{aligned}$$

woraus durch Addition des additiven Inversen von $0\cdot x$ auf beiden Seiten die gewünschte Gleichung $0=0\cdot x$ folgt.

(b) Es ist

$$\begin{aligned} x\cdot(-y)+xy &= x\cdot(-y+y) && \text{(Distributivität)} \\ &= x\cdot 0 && (-y \text{ ist additives Inverses zu } y) \\ &= 0 && \text{(nach (a))}, \end{aligned}$$

daher ist $x\cdot(-y)$ das additive Inverse zu xy, d. h. es gilt $x\cdot(-y) = -(xy)$.

(c) Doppeltes Anwenden von (b), einmal für den linken und einmal für den rechten Faktor, ergibt

$$(-(x^{-1}))\cdot(-x) = -(x^{-1}\cdot(-x)) = -(-(x^{-1}\cdot x)) = -(-1) \overset{3.7(a)}{=} 1.$$

Also ist $-(x^{-1})$ das multiplikative Inverse zu $-x$, d. h. es ist $-(x^{-1}) = (-x)^{-1}$. □

Notation 3.9. In einem Körper K verwendet man üblicherweise die folgenden Notationen, von denen euch die meisten sicher bekannt sein werden:

(a) Für $x,y \in K$ setzt man $x-y := x+(-y)$. Ist $y \neq 0$, so setzt man $\frac{x}{y} := x\cdot y^{-1}$.

(b) Für $x \in K$ und $n \in \mathbb{N}$ definiert man die n-te **Potenz** von x als

$$x^n := \underbrace{x\cdot \cdots \cdot x}_{n\text{-mal}},$$

wobei dieser Ausdruck für $n = 0$ als $x^0 := 1$ zu verstehen ist. Insbesondere legen wir also auch $0^0 := 1$ fest. Beachte, dass aus dieser Definition (und der Kommutativität der Multiplikation) unmittelbar die Potenzrechenregeln

$$x^m\cdot x^n = x^{m+n} \quad \text{und} \quad (xy)^n = x^n\cdot y^n$$

für alle $x,y \in K$ folgen. Ist $x \neq 0$, so definiert man zusätzlich Potenzen mit negativen ganzzahligen Exponenten durch $x^{-n} := (x^{-1})^n$.

Beachte, dass auch in einem beliebigen Körper K die Exponenten einer Potenz stets *ganze Zahlen* sind und keine Elemente aus K. Eine Potenz x^y für $x,y \in K$ lässt sich im Allgemeinen nicht definieren (auch wenn dies für $K = \mathbb{R}$ in vielen Fällen möglich ist, siehe Definition **??**).

(c) Manchmal möchte man mehrere Elemente $x_m, x_{m+1}, x_{m+2}, \dots, x_n$ in einem Körper (oder allgemeiner in einer additiv geschriebenen abelschen Gruppe) aufsummieren, die durch eine ganzzahlige Laufvariable indiziert werden, die von einem $m \in \mathbb{Z}$ bis zu einem $n \in \mathbb{Z}$ (mit $n \geq m$) läuft. Man schreibt dies dann als

$$\sum_{i=m}^{n} x_i := x_m + x_{m+1} + x_{m+2} + \cdots + x_n$$

(also mit einem großen griechischen Sigma, das an das Wort „Summe" erinnern soll). So steht z. B.

$$\sum_{i=1}^{n} i^2 = 1^2+2^2+3^2+\cdots+n^2 \qquad (*)$$

für die Summe aller Quadratzahlen bis n^2. Natürlich ist der Name der Laufvariablen dabei egal, und der Ausdruck $(*)$ hängt nicht von einem i ab (wie man auf der rechten Seite ja auch sieht). Außerdem kann man die Laufvariable verschieben, ohne den eigentlichen Ausdruck zu ändern: Setzt man z. B. $i = j+1$, also $j = i-1$, in der obigen Summe $(*)$, so läuft j dort von 0 bis $n-1$, wenn i von 1 bis n läuft, und wir können dieselbe Summe auch schreiben als

$$\sum_{j=0}^{n-1} (j+1)^2 = 1^2+2^2+3^2+\cdots+n^2.$$

Natürlich kann man diesen Ausdruck nun auch wieder genauso gut mit dem Buchstaben i statt j als $\sum_{i=0}^{n-1}(i+1)^2$ schreiben, oder den Index um mehr als 1 in die eine oder andere Richtung verschieben. Also:

Der Wert einer Summe ändert sich nicht, wenn man zur Laufvariablen im zu summierenden Ausdruck eine Konstante addiert, und dafür von der Ober- und Untergrenze der Summe diese Konstante abzieht.

Wir sagen in diesem Fall, dass die neue Darstellung der Summe durch eine **Indexverschiebung** (im Beispiel oben $i \mapsto i+1$) aus der alten hervorgeht.

Analog schreibt man

$$\prod_{i=m}^{n} x_i := x_m \cdot x_{m+1} \cdot x_{m+2} \cdot \,\cdots\, \cdot x_n$$

(mit einem großen griechischen Pi für das Produkt), wenn man die Körperelemente multiplizieren statt addieren möchte. Ist schließlich die Obergrenze einer Summe oder eines Produkts kleiner als die Untergrenze (man spricht dann von der **leeren Summe** bzw. dem **leeren Produkt**), so definiert man dies als

$$\sum_{i=m}^{n} x_i := 0 \qquad \text{und} \qquad \prod_{i=m}^{n} x_i := 1 \qquad \text{für } n < m,$$

also als das additive bzw. multiplikative neutrale Element.

(d) Ist n eine natürliche Zahl, so fasst man diese oft auch als das Element

$$\sum_{i=1}^{n} 1 = \underbrace{1 + \cdots + 1}_{n\text{-mal}}$$

von K auf. Im Fall $K = \mathbb{R}$ ist dies dann einfach die natürliche Zahl $n \in \mathbb{N} \subset \mathbb{R}$ und liefert somit keine neue Notation, aber z. B. in $K = \mathbb{Z}_2$ aus Beispiel 3.6 (b) ist $2 = 1 + 1 = 0$.

Aufgabe 3.10. Zeige, dass in jedem Körper K die üblichen Rechenregeln

$$\frac{x}{y} + \frac{z}{w} = \frac{xw + yz}{yw} \qquad \text{und} \qquad \frac{x}{y} \cdot \frac{z}{w} = \frac{xz}{yw}$$

für Brüche gelten, wobei $x, y, z, w \in K$ mit $y, w \neq 0$.

Aufgabe 3.11. Es sei $a \in \mathbb{Q}$ fest gegeben. Wir definieren auf $\mathbb{Q}^2$ eine Addition und Multiplikation durch

$$(x_1, x_2) + (y_1, y_2) := (x_1 + y_1, x_2 + y_2) \quad \text{und} \quad (x_1, x_2) \cdot (y_1, y_2) := (x_1 y_1 + a x_2 y_2, x_1 y_2 + x_2 y_1).$$

Man prüft leicht durch explizite Rechnung nach, dass $\mathbb{Q}^2$ mit dieser Addition eine kommutative Gruppe mit neutralem Element $(0,0)$ ist, dass auch die Multiplikation kommutativ ist, und dass diese beiden Operationen das Distributivgesetz erfüllen – dies soll in dieser Aufgabe nicht bewiesen werden. Man zeige stattdessen:

(a) Die Multiplikation ist assoziativ und besitzt ein neutrales Element.

(b) $(\mathbb{Q}^2, +, \cdot)$ ist genau dann ein Körper, wenn a kein Quadrat in $\mathbb{Q}$ ist, also wenn es kein $b \in \mathbb{Q}$ gibt mit $a = b^2$.

Aufgabe 3.12. Zu einem Körper K und einer Menge D mit $|D| \geq 2$ sei

$$V = \{f : f \text{ ist eine Abbildung von } D \text{ nach } K\}$$

die Menge aller reellwertigen Funktionen auf D. Für $f, g \in V$ definieren wir die Addition $f + g$ und Multiplikation $f \cdot g$ dieser Funktionen punktweise durch

$$f + g \colon D \to K,\ x \mapsto f(x) + g(x) \qquad \text{und} \qquad f \cdot g \colon D \to K,\ x \mapsto f(x) \cdot g(x).$$

(a) Zeige, dass V mit dieser Addition eine abelsche Gruppe ist.

(b) Ist V mit dieser Addition und Multiplikation ein Körper?

3.B Vollständige Induktion

Häufig möchte man in der Mathematik Aussagen beweisen, die von einer natürlichen Zahl abhängen – z. B. bei Formeln, die Summen oder Produkte wie in Notation 3.9 mit variablen Unter- oder Obergrenzen beinhalten. Die einfachste und bekannteste solcher Aussagen ist vermutlich die folgende Formel für die Summe aller natürlichen Zahlen bis zu einer gegebenen Obergrenze.

Satz 3.13 (Summenformel von Gauß). *Für alle $n \in \mathbb{N}$ gilt*

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

Beispiel 3.14. Für $n = 5$ ist z. B.

$$\sum_{k=1}^{5} k = 1+2+3+4+5 = 15 = \frac{5 \cdot 6}{2}.$$

Um derartige Aussagen zu beweisen, ist oft das Beweisverfahren der **(vollständigen) Induktion** nützlich, das wir jetzt einführen wollen.

Angenommen, wir wollen (wie z. B. in Satz 3.13) eine Aussage $A(n)$ für alle $n \in \mathbb{N}$ beweisen. Dann können wir dies tun, indem wir die folgenden beiden Dinge zeigen:

(a) (**Induktionsanfang**) Die Aussage $A(0)$ ist wahr.

(b) (**Induktionsschritt**) Für alle $n \in \mathbb{N}$ gilt $A(n) \Rightarrow A(n+1)$, d. h. *wenn* die Aussage $A(n)$ für ein gegebenes $n \in \mathbb{N}$ gilt (die „Induktionsannahme“ bzw. „Induktionsvoraussetzung“), *dann* gilt auch die Aussage $A(n+1)$ (der „Induktionsschluss“).

Haben wir diese beiden Dinge gezeigt, so folgt daraus nämlich die Gültigkeit von $A(n)$ für alle $n \in \mathbb{N}$: Die Aussage $A(0)$ haben wir mit dem Induktionsanfang gezeigt, und durch fortgesetztes Anwenden des Induktionsschritts $A(n) \Rightarrow A(n+1)$ für $n = 0, 1, 2, \dots$ erhalten wir dann auch

$$A(0) \Rightarrow A(1) \Rightarrow A(2) \Rightarrow A(3) \Rightarrow \cdots,$$

also die Gültigkeit von $A(n)$ für alle $n \in \mathbb{N} = \{0, 1, 2, 3, \dots\}$.

Derartige Induktionsbeweise sind immer dann sinnvoll, wenn die Aussagen $A(n)$ und $A(n+1)$ „ähnlich genug“ sind, so dass es beim Beweis von $A(n+1)$ hilft, die Gültigkeit von $A(n)$ voraussetzen zu dürfen.

Mit diesem Verfahren können wir nun die Summenformel aus Satz 3.13 beweisen:

Beweis von Satz 3.13. Wir zeigen die Formel mit Induktion über n.

Induktionsanfang ($n = 0$): Für $n = 0$ stimmen die beiden Seiten der zu zeigenden Gleichung überein, denn es ist

$$\sum_{k=1}^{0} k = 0 = \frac{0 \cdot (0+1)}{2}.$$

Induktionsschritt ($n \to n+1$): Als Induktionsvoraussetzung nehmen wir an, dass die zu beweisende Formel für ein gegebenes $n \in \mathbb{N}$ richtig ist, d. h. dass

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

gilt. (Beachte, dass wir diese Gleichung nicht für alle $n \in \mathbb{N}$ voraussetzen – dies wäre ja schon die gesamte zu zeigende Aussage!) Wir müssen zeigen, dass die entsprechende Gleichung dann auch für $n+1$ gilt, also dass

$$\sum_{k=1}^{n+1} k = \frac{(n+1)(n+2)}{2}.$$

Dies ergibt sich nun leicht aus der folgenden Rechnung:

$$\begin{aligned}
\sum_{k=1}^{n+1} k &= \sum_{k=1}^{n} k + (n+1) && \text{(Abspalten des letzten Summanden für } k=n+1) \\
&= \frac{n(n+1)}{2} + (n+1) && \text{(nach Induktionsvoraussetzung)} \\
&= \frac{n^2+n+2n+2}{2} \\
&= \frac{(n+1)(n+2)}{2}.
\end{aligned}$$

Damit ist der Satz mit vollständiger Induktion bewiesen. □

Bemerkung 3.15. Offensichtlich erlaubt das Beweisverfahren der vollständigen Induktion die folgenden Abwandlungen:

(a) Im Induktionsschritt kann man, wenn es hilfreich ist, beim Beweis der Aussage $A(n+1)$ nicht nur die direkt vorangegangene Aussage $A(n)$, sondern *alle bereits gezeigten Aussagen* $A(0), A(1), \dots, A(n)$ voraussetzen.

(b) Möchte man die Aussage $A(n)$ nicht für alle $n \in \mathbb{N}$, sondern für alle $n \in \mathbb{Z}$ ab einem gewissen Startwert $n_0 \in \mathbb{Z}$ zeigen, so kann man als Induktionsanfang die Aussage $A(n_0)$ zeigen, und im Induktionsschritt dann die Folgerung $A(n) \Rightarrow A(n+1)$ für alle $n \geq n_0$.

Aufgabe 3.16. Zeige für alle $n \in \mathbb{N}$:

$$\text{(a)} \quad \sum_{k=1}^{n} \frac{1}{k(k+1)} = \frac{n}{n+1}, \qquad \text{(b)} \quad \prod_{k=1}^{n} \left(1 + \frac{1}{n+k}\right) = 2 - \frac{1}{n+1}.$$

3.C Polynomfunktionen

Als erste Anwendung der Körpereigenschaften wollen wir zum Abschluss dieses Kapitels die euch sicher schon aus der Schule bekannten Polynomfunktionen behandeln – also die Funktionen, die sich aus den grundlegenden Körperoperationen Addition und Multiplikation bilden lassen.

Definition 3.17 (Polynomfunktionen und Nullstellen)**.** Es seien D eine Teilmenge eines Körpers K und $f : D \to K$ eine Funktion.

(a) Ist f von der Form

$$f(x) = \sum_{k=0}^{n} a_k x^k = a_n x^n + \cdots + a_1 x + a_0$$

für gewisse $a_0, \dots, a_n \in K$, so sagt man, dass f eine **Polynomfunktion** ist. Ist n dabei so gewählt, dass der erste Koeffizient a_n ungleich Null ist, so heißt f eine Polynomfunktion vom **Grad** n und mit **Leitkoeffizient** a_n. Ist der Leitkoeffizient 1, so heißt f eine **normierte** Polynomfunktion.

Sind in der obigen Darstellung alle Koeffizienten $a_0, \dots, a_n$ gleich 0 (und ist f damit die Nullfunktion), so nennen wir f formal eine Polynomfunktion vom Grad $-\infty$. In diesem Fall hat f keinen Leitkoeffizienten.

(b) Ist $x_0 \in D$ mit $f(x_0) = 0$, so nennt man x_0 eine **Nullstelle** von f.

Das Besondere an Nullstellen von Polynomfunktionen ist, dass man sie wie im folgenden Satz als Linearfaktoren abspalten kann.

Satz 3.18 (Abspalten von Nullstellen in Polynomfunktionen)**.** *Es seien K ein Körper, $D \subset K$ und $f : D \to K$ eine Polynomfunktion vom Grad $n \in \mathbb{N}$.*

(a) *Ist $x_0 \in D$ eine Nullstelle von f, so gibt es eine Polynomfunktion $g : D \to K$ vom Grad $n-1$ mit $f(x) = (x-x_0)\,g(x)$ für alle $x \in D$ (d. h. man kann „den Linearfaktor $x - x_0$ abspalten“).*

(b) *Die Funktion f hat höchstens n Nullstellen.*

Beweis. Wir zeigen die beiden Aussagen mit Induktion über n. Der Beweis von (a) ist dabei konstruktiv, d. h. er gibt auch ein Verfahren an, wie g berechnet werden kann (siehe Beispiel 3.19).

Der Induktionsanfang für $n = 0$ ist trivial, denn f ist dann eine Konstante ungleich 0 und hat somit keine Nullstellen. Für den Induktionsschluss nehmen wir an, dass die Aussagen des Satzes bis zu einem gegebenen n gelten, und betrachten $f\colon D \to K$, $x \mapsto a_{n+1}x^{n+1} + \cdots + a_1x + a_0$ vom Grad $n+1$, also mit $a_{n+1} \neq 0$.

(a) Wir definieren eine Polynomfunktion $\tilde{f}\colon D \to K$ durch

$$\begin{aligned}\tilde{f}(x) &:= f(x) - a_{n+1}x^n(x - x_0)\\ &= a_{n+1}x_0x^n + a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0\end{aligned}$$

für alle x. Ist $\tilde{f}$ die Nullfunktion, so sind wir fertig, da dann ja $f(x) = a_{n+1}x^n(x - x_0)$ für alle $x \in D$ gilt. Andernfalls ist $\tilde{f}$ nach Konstruktion eine Polynomfunktion von einem Grad kleiner als $n+1$ (der x^{n+1}-Term hebt sich ja gerade heraus), die immer noch die Nullstelle x_0 hat. Nach Induktionsvoraussetzung gibt es dann also eine Polynomfunktion $\tilde{g}\colon D \to K$ vom Grad kleiner als n mit $\tilde{f}(x) = (x - x_0)\,\tilde{g}(x)$ für alle $x \in D$, und somit ist

$$\begin{aligned}f(x) &= a_{n+1}x^n(x - x_0) + \tilde{f}(x)\\ &= (x - x_0) \cdot \underbrace{(a_{n+1}x^n + \tilde{g}(x))}_{=:g(x)}\end{aligned}$$

für alle $x \in D$, wobei g offensichtlich vom Grad n ist.

(b) Hat f keine Nullstelle, so sind wir fertig. Andernfalls wählen wir eine Nullstelle x_0 von f und schreiben $f(x) = (x - x_0)\,g(x)$ für alle $x \in D$ wie in (a) mit einer Polynomfunktion g vom Grad n. Nach Induktionsvoraussetzung hat g höchstens n Nullstellen, und nach Bemerkung 3.7 (b) sind die Nullstellen von f genau x_0 zusammen mit den Nullstellen von g. Also hat f höchstens $n+1$ Nullstellen. Damit ist die Behauptung mit Induktion bewiesen. □

Beispiel 3.19 (Polynomdivision)**.** Das Verfahren aus dem Beweis von Satz 3.18 (a) wird als *Polynomdivision* [G, Satz 10.19] bezeichnet: Man subtrahiert fortlaufend geeignete Vielfache von $x - x_0$ von f, so dass sich der jeweils höchste Term von f weghebt, und sammelt die dabei verwendeten Faktoren in g. Das folgende Schema, das genauso aussieht wie eine normale schriftliche Division, verdeutlicht dieses Verfahren am Beispiel der Funktion $f\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2 + 3x - 4$ mit Nullstelle $x_0 = 1$, die wir als $f(x) = (x-1)\,g(x)$ für alle $x \in \mathbb{R}$ schreiben wollen. Das Ergebnis ist in diesem Fall $g\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto x + 4$.

$$\begin{array}{lrl}
f \longrightarrow & (x^2 + 3x - 4) & : (x-1) = x + 4 \longleftarrow g\\
& -\ (x^2 - x) & \quad \cdot x\\
\hline
\tilde{f} \longrightarrow & 4x - 4 & \\
& -\ (4x - 4) & \quad \cdot 4\\
\hline
& 0 &
\end{array}$$

Bemerkung 3.20. Satz 3.18 liefert uns zwar die neue Funktion nach dem Abspalten des Linearfaktors, er sagt uns hingegen nicht, wie wir überhaupt erst einmal eine Nullstelle von f finden können, oder ob es überhaupt Nullstellen gibt (die reelle Polynomfunktion $f(x) = x^2 + 1$ hat ja z. B. keine Nullstellen). In der Tat gibt es im Allgemeinen kein Verfahren, wie man Nullstellen von Polynomfunktionen exakt berechnen kann! Genauer gesagt gilt:

- Für Polynomfunktionen vom Grad höchstens 4 gibt es explizite Verfahren zur exakten Bestimmung der Nullstellen (für Grad 1 ist das klar, für Grad 2 gibt es die bekannte „*p*-*q*-Formel“ bzw. die quadratische Ergänzung, und für Grad 3 bzw. 4 sind die Formeln so lang, dass man im Allgemeinen nicht mehr mit ihnen arbeiten möchte).

- Für Polynomfunktionen vom Grad größer als 4 kann man beweisen(!), dass es keine derartigen Verfahren zur exakten Bestimmung der Nullstellen geben kann (das beweist man z. B. in der Vorlesung „Einführung in die Algebra“, die ihr im nächsten Studienjahr hören könnt). Aber:
- Für reelle Polynomfunktionen beliebigen Grades gibt es zumindest numerische Verfahren, die die Nullstellen (mit beliebiger Genauigkeit) näherungsweise bestimmen können.

Zum Schluss wollen wir nun noch zwei wichtige Konzepte für Polynomfunktionen untersuchen, die ihr beide im reellen Fall vielleicht schon aus der Schule kennt: den sogenannten Koeffizientenvergleich (also dass eine Polynomfunktion eindeutig ihre Koeffizienten bestimmt) und die Vielfachheit von Nullstellen. Es stellt sich jedoch heraus, dass man hierfür im allgemeinen Fall die Voraussetzung benötigt, dass die Definitionsmenge der betrachteten Funktionen unendlich viele Elemente besitzt.

Lemma 3.21 (**Koeffizientenvergleich**). *Es seien K ein Körper, $D \subset K$ mit $|D| = \infty$, und $f\colon D \to K$ eine Polynomfunktion mit zwei Darstellungen*

$$f(x) = a_n x^n + \cdots + a_1 x + a_0 = b_n x^n + \cdots + b_1 x + b_0 \quad \text{für alle } x \in D$$

für gewisse $a_0, \ldots, a_n, b_0, \ldots, b_n \in K$. (Beachte, dass wir dabei in beiden Darstellungen den gleichen höchsten Exponenten n wählen können, da wir nicht $a_n \neq 0$ und $b_n \neq 0$ vorausgesetzt haben.)

Dann gilt bereits $a_i = b_i$ für alle $i = 0, \ldots, n$. Es ist also nicht möglich, „eine Polynomfunktion auf zwei verschiedene Arten hinzuschreiben“.

Beweis. Nach Voraussetzung ist die Polynomfunktion

$$D \to K,\ x \mapsto (a_n - b_n)x^n + \cdots + (a_1 - b_1)x + (a_0 - b_0) = f(x) - f(x) = 0$$

die Nullfunktion auf D. Da sie damit wegen $|D| = \infty$ unendlich viele Nullstellen besitzt, muss sie nach Satz 3.18 (b) vom Grad $-\infty$ sein. Also sind alle Koeffizienten dieser Polynomfunktion gleich 0, d. h. es ist $a_i = b_i$ für alle $i = 0, \ldots, n$. □

Bemerkung und Notation 3.22 (Polynome). Die Voraussetzung $|D| = \infty$ in Lemma 3.21 ist wirklich notwendig: So sind für $D = K = \mathbb{Z}_2$ wie in Beispiel 3.6 (b) z. B. $x \mapsto x$ und $x \mapsto x^2$ dieselbe Funktion, da sie beide 0 auf 0 und 1 auf 1 abbilden und in $\mathbb{Z}_2$ keine weiteren Elemente existieren.

In der Literatur bezeichnet man einen formalen Ausdruck der Form $a_n x^n + \cdots + a_1 x + a_0$ mit $a_0, \ldots, a_n \in K$ als ein **Polynom** über K [G, Kapitel 9]. Jedes solche Polynom bestimmt natürlich eine Polynomfunktion von jeder Teilmenge D von K nach K, allerdings können verschiedene Polynome wie im eben angegebenen Beispiel durchaus dieselbe Polynomfunktion definieren: Über $\mathbb{Z}_2$ sind x und x^2 verschiedene Polynome, sie bestimmen aber dieselbe Polynomfunktion.

Mit dieser Notation ist die Aussage von Lemma 3.21 also, dass Polynome und Polynomfunktionen im Fall von unendlichen Definitionsmengen dasselbe sind. Da wir Polynomfunktionen im Folgenden in der Regel nur in diesem Fall unendlicher Definitionsmengen benötigen, werden wir die Begriffe Polynom und Polynomfunktion oft synonym verwenden. Wegen der Eindeutigkeit der Koeffizienten sind dann auch der Grad (und der Leitkoeffizient) einer Polynomfunktion f wie in Definition 3.17 (a) eindeutig bestimmt. Wir können daher eine Bezeichnung dafür einführen:

Definition 3.23 (Grad eines Polynoms). Wir bezeichnen den **Grad** einer Polynomfunktion f (mit unendlicher Definitionsmenge) mit $\deg f \in \mathbb{N} \cup \{-\infty\}$ (vom englischen Wort „degree“).

Satz und Definition 3.24 (Vielfachheit von Nullstellen). *Es seien K ein Körper, $D \subset K$ mit $|D| = \infty$ und $f\colon D \to K$ eine Polynomfunktion, die nicht die Nullfunktion ist. Dann lässt sich f (bis auf die Reihenfolge der Faktoren) eindeutig als*

$$f(x) = g(x) \cdot (x - x_1)^{a_1} \cdot \cdots \cdot (x - x_k)^{a_k} \qquad \text{für alle } x \in D$$

schreiben, wobei $x_1, \ldots, x_k \in D$ die verschiedenen Nullstellen von f sind, $a_1, \ldots, a_k \in \mathbb{N}_{>0}$ gilt, und g eine Polynomfunktion ohne Nullstellen in D ist. In dieser Darstellung nennt man a_i für $i = 1, \ldots, k$ die ***Vielfachheit*** *der Nullstelle x_i (in der Literatur sind auch die Bezeichnungen* ***Ordnung*** *und* ***Multiplizität*** *der Nullstelle üblich).*

Beweis. Die Existenz einer solchen Darstellung ergibt sich sofort durch fortgesetztes Abspalten von Nullstellen gemäß Satz 3.18 (a). Wir zeigen nun die Eindeutigkeit mit Induktion über den Grad der Polynomfunktion. Dabei ist der Induktionsanfang für Grad 0 trivial, denn dann hat f keine Nullstellen, und es ist zwangsläufig $k = 0$ und $g = f$.

Für den Induktionsschritt bemerken wir zuerst, dass $x_1, \dots, x_k$ natürlich in jedem Fall als die Nullstellen von f eindeutig bestimmt sind. Wir nehmen also an, dass wir zwei Darstellungen

$$f(x) = g(x) \cdot (x-x_1)^{a_1} \cdot \dots \cdot (x-x_k)^{a_k} = h(x) \cdot (x-x_1)^{b_1} \cdot \dots \cdot (x-x_k)^{b_k}$$

wie in der Behauptung des Satzes haben. Im nullstellenfreien Fall $k = 0$ sind wir natürlich bereits fertig. Andernfalls liefert Division durch $x - x_1$ für alle $x \in D \backslash \{x_1\}$ (wir müssen x_1 hier herausnehmen, da wir sonst durch 0 teilen würden!)

$$\begin{aligned} &g(x) \cdot (x-x_1)^{a_1-1} \cdot (x-x_2)^{a_2} \cdot \dots \cdot (x-x_k)^{a_k} \\ &= h(x) \cdot (x-x_1)^{b_1-1} \cdot (x-x_2)^{b_2} \cdot \dots \cdot (x-x_k)^{b_k}. \end{aligned} \tag{$*$}$$

Wir haben also wieder zwei Darstellungen einer Polynomfunktion auf der immer noch unendlichen Menge $D \backslash \{x_1\}$. Da der Grad dieser Polynomfunktion nun um 1 kleiner ist als der von f, müssen diese Darstellungen aber nach der Induktionsvoraussetzung bereits übereinstimmen. Also gilt $g = h$, $a_1 - 1 = b_1 - 1$, $a_2 = b_2$, ..., $a_k = b_k$, und damit stimmen auch die beiden ursprünglichen Darstellungen von f überein. □

Aufgabe 3.25. Bestimme die Nullstellen des reellen Polynoms $x \mapsto x^4 + 3x^3 - 4x$ und ihre Vielfachheiten.

13. Lineare Gleichungssysteme und Matrizen

Ausgehend von den elementaren Konzepten in den Kapiteln 1 bis 3 wollen wir in dieser Vorlesung zwei grundlegende Gebiete der Mathematik entwickeln: die Analysis und die lineare Algebra. Während sich die eindimensionale Analysis in den Kapiteln **??** bis **??** dabei hauptsächlich mit allgemeinen (in der Regel stetigen oder sogar differenzierbaren) Funktionen in einer reellen Variablen beschäftigt hat, wollen wir nun unabhängig davon im folgenden Teil des Skripts (bis Kapitel **??**) zunächst *lineare* Funktionen in *mehreren* Variablen studieren. Später werden wir die erarbeiteten Resultate dann mit der Analysis kombinieren, um auch Funktionen in mehreren Variablen mit Hilfe von Ableitungen linear approximieren zu können.

In der Analysis arbeitet man ja hauptsächlich über den reellen Zahlen, und das ist in der linearen Algebra auch nicht viel anders. Allerdings haben wir in Kapitel 3 ja auch schon gesehen, dass viele Dinge auch über einem beliebigen Körper funktionieren. Die lineare Algebra verhält sich hier sehr „gutartig“: Da wir letztlich nur lineare Funktionen betrachten werden, benötigen wir gar nicht mehr Struktur der reellen Zahlen als die Körperaxiome. Wir können daher nahezu die gesamte lineare Algebra über einem beliebigen Grundkörper studieren, also z. B. auch über $\mathbb{Q}$, dem Körper $\mathbb{Z}_2$ aus Beispiel 3.6 (b), oder anderen Körpern, die ihr vielleicht inzwischen kennt. Wir vereinbaren daher:

Im Folgenden (bis Kapitel **??**) sei K immer ein fest gewählter Grundkörper.

Beim ersten Lesen könnt ihr euch K aber durchaus gerne einfach als $\mathbb{R}$ vorstellen.

Wenn wir uns nun eine ganze Weile lang nur mit linearen Funktionen bzw. Gleichungen beschäftigen, werden nahezu alle unsere Konzepte und Probleme auf rechnerischer Seite am Ende auf *lineare Gleichungssysteme* der Form

$$\begin{aligned} a_{1,1}x_1 + \cdots + a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + \cdots + a_{2,n}x_n &= b_2 \\ &\;\vdots \\ a_{m,1}x_1 + \cdots + a_{m,n}x_n &= b_m \end{aligned} \qquad (*)$$

hinauslaufen, d. h. zu gegebenen $m,n \in \mathbb{N}$ sowie $a_{i,j}, b_i \in K$ für $i = 1,\dots,m$ und $j = 1,\dots,n$ wollen wir alle $(x_1,\dots,x_n) \in K^n$ bestimmen, die diese Gleichungen $(*)$ simultan erfüllen. Um diese Probleme dann auch gleich lösen zu können, wollen wir daher in diesem und dem folgenden Kapitel zunächst einmal studieren, wie die Lösungsmengen solcher linearen Gleichungssysteme aussehen und konkret berechnet werden können. An manchen Stellen können wir dabei auch schon Ausblicke auf die abstrakten Konzepte geben, die wir später in der Vorlesung einführen werden und die der eigentliche Gegenstand der linearen Algebra sind (wie z. B. Vektorräume, lineare Abbildungen, lineare Unabhängigkeit oder Dimension).

13.A Vektoren und Matrizen

Bevor wir mit der Untersuchung linearer Gleichungssysteme beginnen, sollten wir als Erstes eine effizientere Notation dafür einführen, da die Schreibweise $(*)$ in der Einleitung oben natürlich sehr unübersichtlich ist und auch viel Platz wegnimmt. Dies ist mit Hilfe von sogenannten *Matrizen* möglich. Die Idee besteht dabei einfach darin, die $m \cdot n$ Zahlen $a_{i,j}$ in $(*)$ in einem mathematischen Objekt zusammenzufassen.

Definition 13.1 (Matrizen)**.** Es seien K ein Körper und $m,n \in \mathbb{N}$.

(a) Eine $m \times n$-**Matrix** mit Einträgen in K ist ein rechteckiges Zahlenschema

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \quad \text{mit } a_{i,j} \in K \text{ für } 1 \le i \le m \text{ und } 1 \le j \le n$$

mit m Zeilen und n Spalten. Analog zur Schreibweise für Zahlenfolgen bezeichnen wir eine solche Matrix auch mit

$$(a_{i,j})_{\substack{1 \le i \le m \\ 1 \le j \le n}} \quad \text{oder einfach} \quad (a_{i,j})_{i,j}.$$

Es ist eine Konvention, dass wir in diesem Fall hinter der Klammer immer zuerst den Zeilenindex und dann den Spaltenindex schreiben (Merkregel: „Zeile zuerst, Spalte später").

Die Menge aller $m \times n$-Matrizen mit Einträgen in K wird mit $K^{m \times n}$ bezeichnet – auch hier steht in der Bezeichnung also zuerst die Anzahl der Zeilen und dann die Anzahl der Spalten.

(b) Die $m \times n$-Matrix, deren Einträge alle 0 sind, heißt **Nullmatrix**. Wir schreiben sie in der Regel einfach als $0 \in K^{m \times n}$.

(c) Eine Matrix mit gleich vielen Zeilen wie Spalten (also $m = n$) heißt **quadratisch**. Für eine solche quadratische Matrix bezeichnet man die Einträge $a_{i,i}$ mit $i = 1, \dots, n$ (also die Einträge von links oben nach rechts unten) als ihre **Diagonaleinträge**.

Notation 13.2 (Vektoren und Skalare)**.** In der linearen Algebra schreibt man die Komponenten eines Elements der Produktmenge K^m *untereinander* als $m \times 1$-Matrix (statt wie sonst üblich nebeneinander) und identifiziert somit K^m mit $K^{m \times 1}$, also mit Matrizen mit nur einer Spalte. Diese Konvention wird sich bei der Einführung von Matrixprodukten in Abschnitt 13.B als nützlich erweisen.

Als vorläufige Notation bis Kapitel **??** bezeichnen wir diese Elemente von $K^m = K^{m \times 1}$ als **Vektoren**, die Elemente von K als **Skalare**. Vektoren sind in diesem Sinne also zunächst Spezialfälle von Matrizen. Den Vektor in K^m, dessen Einträge alle 0 sind, nennt man dementsprechend wie in Definition 13.1 (b) auch **Nullvektor**.

Oft kommt auch der Vektor in K^m vor, dessen Einträge nur in einer Zeile i gleich 1 und sonst überall gleich 0 sind. Man nennt ihn den i-ten **Einheitsvektor** und schreibt ihn als e_i. Die Einheitsvektoren in K^3 sind also z. B.

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{und} \quad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Definition 13.3 (Matrixoperationen)**.** Für zwei Matrizen $A = (a_{i,j})_{i,j}$ und $B = (b_{i,j})_{i,j}$ in $K^{m \times n}$ sowie $\lambda \in K$ definieren wir

(a) die Addition $A + B := (a_{i,j} + b_{i,j})_{i,j} \in K^{m \times n}$;

(b) die Skalarmultiplikation $\lambda A := \lambda \cdot A := (\lambda \cdot a_{i,j})_{i,j} \in K^{m \times n}$;

(c) die **transponierte Matrix** $A^{\mathsf{T}} := (a_{j,i})_{i,j} \in K^{n \times m}$.

Beachte, dass wir dabei die Addition und Skalarmultiplikation von Matrizen mit den gleichen Symbolen „$+$" bzw. „$\cdot$" wie die entsprechenden Verknüpfungen in K bezeichnet haben. Dies wird im Folgenden nicht zu Verwirrungen führen, da wir durch die Art der verknüpften Objekte immer sofort erkennen können, ob es sich um die Rechenoperationen in K oder die oben Operationen für Matrizen handelt.

Beispiel 13.4.

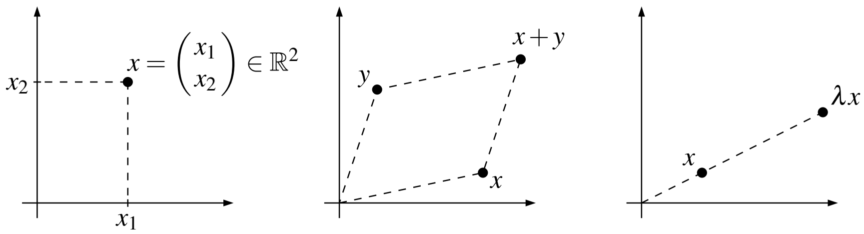
(a) Die Addition und Skalarmultiplikation für Matrizen (und damit auch für Vektoren) sind einfach komponentenweise definiert, es ist also z. B.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 4 \\ 6 & 7 & 8 \end{pmatrix} \in K^{2 \times 3} \quad \text{und} \quad 2 \cdot \begin{pmatrix} 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \end{pmatrix} \in K^2.$$

Die Transposition hingegen vertauscht die Rolle von Zeilen und Spalten in der Matrix, wie z. B. in

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^{\mathsf{T}} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \in K^{3\times 2}.$$

(b) Im Fall $K = \mathbb{R}$ und $n = 2$ ist $K^n = \mathbb{R}^2$ einfach die bekannte reelle Ebene, und Addition und Skalarmultiplikation können wie im folgenden Bild veranschaulicht werden. Die geometrische Interpretation im Fall von $\mathbb{R}^n$ für andere n ist natürlich analog.



Lemma 13.5 (Eigenschaften der Addition und Skalarmultiplikation von Matrizen)**.** *Für alle natürlichen Zahlen m,n, Matrizen $A,B \in K^{m\times n}$ und Skalare $\lambda,\mu \in K$ gilt:*

(a) $(K^{m\times n},+)$ *ist eine abelsche Gruppe (siehe Definition 3.1).*

(b) *(1. Distributivität)* $(\lambda+\mu)A = \lambda A + \mu A$.

(c) *(2. Distributivität)* $\lambda(A+B) = \lambda A + \lambda B$.

(d) *(Assoziativität)* $(\lambda\mu)A = \lambda(\mu A)$.

(e) $1 \cdot A = A$.

Beweis. Die Beweise folgen alle aus den entsprechenden Eigenschaften von K und ergeben sich durch einfaches Nachrechnen. Wir zeigen hier exemplarisch die 1. Distributivität (b): Mit $A = (a_{i,j})_{i,j}$ gilt

$$\begin{aligned}
(\lambda+\mu)\cdot A &= \big((\lambda+\mu)a_{i,j}\big)_{i,j} && \text{(Definition der Skalarmultiplikation)}\\
&= \big(\lambda a_{i,j} + \mu a_{i,j}\big)_{i,j} && \text{(Distributivität in } K\text{)}\\
&= \big(\lambda a_{i,j}\big)_{i,j} + \big(\mu a_{i,j}\big)_{i,j} && \text{(Definition der Matrixaddition)}\\
&= \lambda A + \mu A. && \text{(Definition der Skalarmultiplikation)}
\end{aligned}$$

Die anderen Eigenschaften zeigt man analog. □

Bemerkung 13.6 (Vektorräume)**.** In der Tat ist der Begriff eines Vektors eigentlich viel allgemeiner als in Notation 13.2 – genau wie bei Gruppen und Körpern in Kapitel 3.A werden allgemeine Vektoren über die Operationen definiert, die man mit ihnen durchführen kann. Dies sind gerade die Addition und Skalarmultiplikation, und die Eigenschaften, die man von diesen Verknüpfungen axiomatisch verlangt, sind genau die aus Lemma 13.5. In diesem Sinne bilden dann also nicht nur die Elemente von K^m einen Vektorraum, sondern auch Matrizen in $K^{m\times n}$, aber auch noch viele ganz andere Objekte wie z. B. Polynome, Funktionen oder Folgen.

Das Studium von Vektorräumen in diesem allgemeinen Sinne ist der eigentliche Inhalt der linearen Algebra. Wir werden damit in Kapitel **??** beginnen. Dort werden wir dann auch schnell sehen, dass sich viele Vektorräume am Ende sehr ähnlich wie K^m verhalten (siehe **??**). Unsere momentane Untersuchung von Vektoren K^m wird daher auch für die allgemeine Theorie von Vektorräumen noch sehr nützlich sein.

Lemma 13.7 (Eigenschaften der Transposition von Matrizen)**.** *Für alle natürlichen Zahlen m,n, Matrizen $A,B \in K^{m\times n}$ und Skalare $\lambda \in K$ gilt*

$$(A+B)^{\mathsf{T}} = A^{\mathsf{T}} + B^{\mathsf{T}} \quad \text{und} \quad (\lambda A)^{\mathsf{T}} = \lambda(A^{\mathsf{T}}).$$

Beweis. Wie Lemma 13.5 ergeben sich auch diese Aussagen durch einfaches Nachrechnen, wir zeigen exemplarisch die erste: Für $A = (a_{i,j})_{i,j}$ und $B = (b_{i,j})_{i,j}$ ist

$$(A+B)^\mathsf{T} = (a_{j,i}+b_{j,i})_{i,j} = (a_{j,i})_{i,j} + (b_{j,i})_{i,j} = A^\mathsf{T} + B^\mathsf{T}. \qquad \square$$

Bemerkung 13.8 (Lineare Abbildungen)**.** Abbildungen wie die Transposition von Matrizen, die im Sinne von Lemma 13.7 mit der Addition und Skalarmultiplikation vertauschen, werden wir später *lineare Abbildungen* nennen und ausführlich untersuchen (siehe **??**). Zusammen mit Vektorräumen wie in Bemerkung 13.6 sind die linearen Abbildungen zwischen ihnen der Hauptgegenstand der linearen Algebra.

13.B Matrixmultiplikation

Die Addition und Skalarmultiplikation von Matrizen waren einfach komponentenweise definiert und haben somit die rechteckige Struktur einer Matrix überhaupt nicht benutzt. Es gibt aber noch eine weitere sehr wichtige Rechenoperation für Matrizen, bei die Form der Matrix eine entscheidende Rolle spielt: die sogenannte Matrixmultiplikation, die wir jetzt einführen wollen.

Definition 13.9 (Matrixmultiplikation). Für natürliche Zahlen m,n,p seien $A = (a_{i,j})_{i,j} \in K^{m\times n}$ und $B = (b_{j,k})_{j,k} \in K^{n\times p}$, d. h. die Matrix B habe so viele Zeilen wie A Spalten. Dann definieren wir das Matrixprodukt AB als

$$AB := A \cdot B := \left(\sum_{j=1}^{n} a_{i,j} b_{j,k} \right)_{i,k} \quad \in K^{m\times p}$$

(merke: Es wird über die „mittleren Indizes“ summiert, also über den Spaltenindex der ersten und den Zeilenindex der zweiten Matrix). Das Produkt AB hat also so viele Zeilen wie die erste Matrix und so viele Spalten wie die zweite.

Beispiel 13.10.

(a) Es ist z. B.

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 1\cdot 5+2\cdot 7 & 1\cdot 6+2\cdot 8 \\ 3\cdot 5+4\cdot 7 & 3\cdot 6+4\cdot 8 \end{pmatrix} = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix}$$

(hier ist $m = n = p = 2$). Im Gegensatz dazu ist das Matrixprodukt

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$$

nicht definiert, weil die erste Matrix nur eine Spalte, die zweite aber zwei Zeilen hat. Der Einfachheit halber werden wir in Zukunft bei einem Matrixprodukt AB stets voraussetzen, dass die zweite Matrix so viele Zeilen hat wie die erste Spalten, und dies nicht jedesmal wieder erwähnen.

(b) Ist $A = (a_{i,j})_{i,j} \in K^{m\times n}$ und $x \in K^n$ mit Komponenten $x_1,\dots,x_n$, so können wir x gemäß Notation 13.2 als Matrix mit nur einer Spalte auffassen und erhalten das Matrixprodukt

$$Ax = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{1,1}x_1 + \cdots + a_{1,n}x_n \\ \vdots \\ a_{m,1}x_1 + \cdots + a_{m,n}x_n \end{pmatrix} \in K^{m\times 1} = K^m. \qquad (*)$$

Wir können also auch eine Matrix mit einem Vektor (der passenden Größe) multiplizieren, um wieder einen Vektor zu erhalten.

(c) Beachte, dass $(*)$ in (b) genau die linke Seite eines linearen Gleichungssystems wie in der Einleitung zu diesem Kapitel ist. Wir können (und werden) lineare Gleichungssysteme in Zukunft also einfach als $Ax = b$ schreiben, wobei $A \in K^{m\times n}$ eine gegebene Matrix, $b \in K^m$ ein gegebener Vektor und $x \in K^n$ der gesuchte Vektor sind.

(d) Ein einfacher, aber oft vorkommender und daher wichtiger Spezialfall von (b) ist, wenn $x = e_j$ für $j = 1, \dots, n$ der j-te Einheitsvektor ist: In diesem Fall ist Ae_j gerade die j-te Spalte von A.

Auch für die Matrixmultiplikation wollen wir zunächst einmal die grundlegenden Rechenregeln angeben bzw. beweisen.

Lemma 13.11 (Eigenschaften der Matrixmultiplikation)**.** *Für alle Matrizen A, B, C passender Größe (d. h. so dass die betrachteten Summen und Produkte definiert sind) sowie $\lambda \in K$ gilt:*

(a) *(Distributivität)* $(A+B)C = AC + BC$ *und* $A(B+C) = AB + AC$.

(b) *(Verträglichkeit mit der Skalarmultiplikation)* $(\lambda A)B = A(\lambda B) = \lambda(AB)$.

(c) *(Assoziativität)* $(AB)C = A(BC)$.

(d) *(Verträglichkeit mit der Transposition)* $(AB)^\mathsf{T} = B^\mathsf{T} A^\mathsf{T}$.

Das Matrixprodukt ist jedoch im Allgemeinen nicht kommutativ (aufgrund der Größenbedingung ist das Produkt AB ja auch nicht einmal genau dann definiert, wenn BA es ist).

Beweis. Auch hier ist der Beweis direktes Nachrechnen. Wir zeigen diesmal exemplarisch die Assoziativität in (c): Für $A = (a_{i,j})_{i,j} \in K^{m\times n}$, $B = (b_{j,k})_{j,k} \in K^{n\times p}$ und $C = (c_{k,l})_{k,l} \in K^{p\times r}$ gilt nach Definition der Matrixmultiplikation

$$AB = \Big(\sum_{j=1}^n a_{i,j} b_{j,k}\Big)_{i,k} \quad \text{und} \quad BC = \Big(\sum_{k=1}^p b_{j,k} c_{k,l}\Big)_{j,l},$$

und damit

$$(AB)C = \Big(\sum_{k=1}^p \Big(\sum_{j=1}^n a_{i,j} b_{j,k}\Big) c_{k,l}\Big)_{i,l} \quad \text{und} \quad A(BC) = \Big(\sum_{j=1}^n a_{i,j} \Big(\sum_{k=1}^p b_{j,k} c_{k,l}\Big)\Big)_{i,l}.$$

Nach der Distributivität in K stimmen diese beiden Ausdrücke aber überein: Durch Ausmultiplizieren ergibt sich in beiden Fällen

$$(AB)C = \Big(\sum_{j=1}^n \sum_{k=1}^p a_{i,j} b_{j,k} c_{k,l}\Big)_{i,l} = A(BC). \qquad \square$$

Aufgabe 13.12 (Blockmatrixmultiplikation). Es seien $A \in K^{m\times n}$ und $B \in K^{n\times p}$ zwei Matrizen, die in „Blockform“

$$A = \left(\begin{array}{c|c} A_1 & A_2 \\ \hline A_3 & A_4 \end{array}\right) \quad \text{bzw.} \quad B = \left(\begin{array}{c|c} B_1 & B_2 \\ \hline B_3 & B_4 \end{array}\right)$$

mit $A_1 \in K^{m_1\times n_1}$ und $B_1 \in K^{n_1\times p_1}$ gegeben sind. Zeige, dass das Matrixprodukt AB dann ebenfalls in Blockform

$$AB = \left(\begin{array}{c|c} A_1B_1 + A_2B_3 & A_1B_2 + A_2B_4 \\ \hline A_3B_1 + A_4B_3 & A_3B_2 + A_4B_4 \end{array}\right)$$

berechnet werden kann (wobei die Blöcke formal genauso multipliziert und addiert werden, als würde man das Produkt zweier Matrizen der Größe 2×2 berechnen). Analog funktioniert diese Rechenregel auch für eine Aufteilung in noch mehr Blöcke.

Bemerkung 13.13 (Matrixmultiplikation als lineare Abbildung)**.** Wie in Beispiel 13.10 (b) liefert jede Matrix $A \in K^{m\times n}$ eine Abbildung

$$f_A \colon K^n \to K^m,\ x \mapsto Ax.$$

Nach Lemma 13.11 vertauscht sie mit der Addition und Skalarmultiplikation, d. h. es gilt

$$f_A(x+y) = A(x+y) = Ax + Ay = f_A(x) + f_A(y) \quad \text{und} \quad f_A(\lambda x) = A(\lambda x) = \lambda(Ax) = \lambda f_A(x)$$

für alle $x, y \in K^n$ und $\lambda \in K$. Wie die Transposition ist sie also eine lineare Abbildung im Sinne von Bemerkung 13.8. In der Tat werden wir in **??** noch sehen, dass jede lineare Abbildung von K^n nach K^m von dieser Form ist und diese linearen Abbildungen so genau den Matrizen in $K^{m\times n}$ entsprechen.

Verkettet man solche Abbildungen miteinander, so ist (für Matrizen geeigneter Größe)

$$(f_A \circ f_B)(x) = A(Bx) \overset{13.11\,(c)}{=} (AB)x = f_{AB}(x),$$

d. h. man kann sich die allgemeine Matrixmultiplikation so vorstellen, dass sie der Verkettung der zugehörigen linearen Abbildungen entspricht.

Nach Lemma 13.11 (c) ist die Matrixmultiplikation assoziativ. Es ist daher naheliegend zu untersuchen, ob sie auch die anderen Gruppenaxiome aus Definition 3.1 erfüllt. Damit sie überhaupt zwischen allen Matrizen definiert ist, sollten wir uns dazu auf quadratische Matrizen beschränken. Ein – wie in Definition 3.1 (b) gefordertes – neutrales Element ist dann schnell gefunden:

Konstruktion 13.14 (Einheitsmatrix)**.** Für $n \in \mathbb{N}$ definieren wir die **Einheitsmatrix** der Größe $n \times n$ als

$$E_n := (\, e_1 \mid \cdots \mid e_n \,) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \in K^{n \times n}.$$

Wir schreiben sie oft auch einfach als E, wenn ihre Größe aus dem Zusammenhang klar ist. In der Literatur ist auch die Bezeichnung I_n oder I üblich. Offensichtlich können wir sie auch schreiben als

$$E_n = (\delta_{i,j})_{i,j} \quad \text{mit} \quad \delta_{i,j} = \begin{cases} 1 & \text{falls } i = j, \\ 0 & \text{falls } i \neq j. \end{cases}$$

Damit sehen wir sofort für jede Matrix $A \in K^{m \times n}$, dass

$$E_m A = \left(\sum_{j=1}^{n} \delta_{i,j} a_{j,k} \right)_{i,k} = (a_{i,k})_{i,k} = A,$$

und analog $AE_n = A$. Die Einheitsmatrizen sind in diesem Sinne also sogar für nicht-quadratische Matrizen neutral für die Matrixmultiplikation.

Inverse Matrizen finden wir jedoch nicht immer. Um dies genauer zu untersuchen, definieren wir die Invertierbarkeit zunächst als Eigenschaft.

Definition 13.15 (Inverse Matrizen)**.** Es seien $n \in \mathbb{N}$ und $A \in K^{n \times n}$. Eine Matrix $A' \in K^{n \times n}$ heißt eine **inverse Matrix** zu A, wenn $A'A = AA' = E_n$. Existiert eine solche zu A inverse Matrix, so nennt man A **invertierbar**.

Die Menge aller invertierbaren Matrizen in $K^{n \times n}$ wird mit $\mathrm{GL}(n,K)$ bezeichnet – der Name kommt von der englischen Bezeichnung „general linear group“ und rührt daher, dass $\mathrm{GL}(n,K)$ in der Tat eine Gruppe ist, wie wir jetzt zeigen werden.

Lemma 13.16 (Invertierbare Matrizen als Gruppe)**.** *Für alle $n \in \mathbb{N}$ ist die Menge $\mathrm{GL}(n,K)$ aller invertierbaren $n \times n$-Matrizen eine Gruppe bezüglich der Matrixmultiplikation.*

Beweis. Wir zeigen zunächst, dass die Matrixmultiplikation wirklich eine Verknüpfung auf der Menge $\mathrm{GL}(n,K)$ ist, also dass für $A, B \in \mathrm{GL}(n,K)$ auch $AB \in \mathrm{GL}(n,K)$ gilt: Sind A' und B' invers zu A bzw. B, so ist

$$B'A' \cdot AB = B'EB = B'B = E \quad \text{und} \quad AB \cdot B'A' = AEA' = AA' = E. \tag{$*$}$$

Also ist $B'A'$ invers zu AB, und damit $AB \in \mathrm{GL}(n,K)$.

Die Gruppenaxiome sind ebenfalls schnell überprüft:

- Die Matrixmultiplikation ist (sogar für beliebige Matrizen) assoziativ nach Lemma 13.11 (c).
- Die Einheitsmatrix $E \in \mathrm{GL}(n,K)$ aus Konstruktion 13.14 ist ein neutrales Element.
- Für $A \in \mathrm{GL}(n,K)$ existiert nach Definition eine inverse Matrix $A' \in K^{n \times n}$ mit $A'A = AA' = E$, und dieselbe Gleichung zeigt auch, dass $A' \in \mathrm{GL}(n,K)$.

Also ist $\mathrm{GL}(n,K)$ mit der Matrixmultiplikation eine Gruppe. □

Folgerung 13.17 (Eigenschaften invertierbarer Matrizen)**.** *Es seien $n \in \mathbb{N}$ und $A,B \in \mathrm{GL}(n,K)$. Dann gilt:*

(a) *Es gibt genau eine zu A inverse Matrix. Wir bezeichnen sie mit A^{-1}.*

(b) $(AB)^{-1} = B^{-1}A^{-1}$.

(c) $(A^{-1})^{-1} = A$.

(d) *Für alle Vektoren $b \in K^n$ hat das lineare Gleichungssystem $Ax = b$ genau eine Lösung, nämlich $x = A^{-1}b$.*

Beweis. Da $\mathrm{GL}(n,K)$ nach Lemma 13.16 eine Gruppe ist, folgen die Eigenschaften (a), (b) und (c) unmittelbar aus Lemma 3.4 (b), (c) bzw. (d). Für (d) müssen wir nur bemerken, dass die Gleichungen $Ax = b$ und $x = A^{-1}b$ durch Multiplikation mit A^{-1} bzw. A zueinander äquivalent sind. □

Für eine invertierbare (quadratische) Matrix A können wir mit Folgerung 13.17 (d) also zumindest theoretisch eine Lösung des Gleichungssystems $Ax = b$ hinschreiben. Es bleiben aber natürlich noch die folgenden beiden Fragen:

- Wie können wir überhaupt bestimmen, ob eine quadratische Matrix A invertierbar ist, und in diesem Fall die inverse Matrix A^{-1} berechnen?
- Wie können wir lineare Gleichungssysteme $Ax = b$ für nicht invertierbare bzw. nicht quadratische Matrizen A lösen?

Beide Fragen werden wir im nächsten Kapitel beantworten. Zum Abschluss dieses Kapitels wollen wir nur noch kurz ein paar Beispiele angeben, in denen wir die Invertierbarkeit bzw. Nichtinvertierbarkeit von Matrizen einfach nachprüfen können.

Beispiel 13.18.

(a) Die Matrix $A = \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}$ ist invertierbar mit $A^{-1} = \begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix}$, denn es ist

$$\begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix} = E_2 \quad \text{und} \quad \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix} = E_2.$$

In der Tat werden wir in **??** noch sehen, dass es genügen würde, nur eine dieser beiden Gleichungen zu überprüfen.

(b) Das lineare Gleichungssystem

$$\begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \end{pmatrix}, \quad \text{also} \quad \begin{aligned} 1x_1 + 3x_2 &= 5 \\ 1x_2 &= 2 \end{aligned}$$

hat nach (a) und Folgerung 13.17 (d) die (natürlich auch direkt offensichtliche) eindeutige Lösung

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 5 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 5 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix},$$

also $x_1 = -1$ und $x_2 = 2$.

(c) Es sei $A \in K^{n \times n}$ eine Matrix, in der in einer Zeile $i \in \{1,\dots,n\}$ alle Einträge gleich 0 sind. Dann sind für jede Matrix $B \in K^{n \times n}$ nach Definition der Matrixmultiplikation auch alle Einträge von AB in Zeile i gleich 0. Insbesondere kann AB damit niemals gleich E, also A nicht invertierbar sein.

Analog ist auch eine quadratische Matrix, bei der in einer Spalte alle Einträge gleich 0 sind, sicher nicht invertierbar.

Aufgabe 13.19. Für ein gegebenes $a \in \mathbb{R}$ sei $A = \begin{pmatrix} a & 2 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{pmatrix} \in \mathbb{R}^{3 \times 3}$.

(a) Für $a \neq 0$ berechne man A^n für alle $n \in \mathbb{N}_{>0}$.

(Hinweis: Bestimme A^n für kleine Werte von n, stelle damit eine Vermutung für die allgemeine Form von A^n auf, und beweise diese Vermutung dann mit vollständiger Induktion.)

(b) Für welche a ist die Matrix A invertierbar? Bestimme in diesen Fällen auch die inverse Matrix A^{-1}.

14. Der Gauß-Algorithmus für lineare Gleichungssysteme

Im letzten Kapitel haben wir lineare Gleichungssysteme

$$\begin{aligned} a_{1,1}x_1 + \cdots + a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + \cdots + a_{2,n}x_n &= b_2 \\ &\vdots \\ a_{m,1}x_1 + \cdots + a_{m,n}x_n &= b_m \end{aligned} \tag{1}$$

über einem gegebenen Körper K betrachtet und sie zunächst einmal in Matrixform als

$$Ax = b \quad \text{mit} \quad A \in K^{m \times n}, x \in K^n \text{ und } b \in K^m \tag{2}$$

umgeschrieben, um leichter mit ihnen umgehen zu können. Wir wollen nun untersuchen, wie die Lösungsmengen solcher Gleichungssysteme aussehen, und dabei auch Algorithmen (d. h. Rechenverfahren) entwickeln, um diese Lösungsmengen auch konkret berechnen zu können. Außerdem werden wir mit diesen Verfahren auch bestimmen können, ob eine gegebene quadratische Matrix im Sinne von Definition 13.15 invertierbar ist, und in diesem Fall ihre inverse Matrix berechnen können.

14.A Elementarmatrizen und Zeilenstufenformen

Die Idee für die Lösung linearer Gleichungssysteme ist sehr naheliegend: Wir wollen die gegebenen Gleichungen in (1) oben so umformen und miteinander kombinieren, dass die neuen Gleichungen im Idealfall nur noch jeweils eine Variable beinhalten und ihre Lösung damit leicht abgelesen werden kann. Dies können wir z. B. tun, indem wir auf geschickte Art mehrfach hintereinander ein geeignetes Vielfaches einer Gleichung zu einer anderen addieren, so dass dabei möglichst viele Variablen herausfallen.

Beachte, dass in der Matrixschreibweise (2) des Gleichungssystems jede Gleichung aus (1) einer Zeile der Matrix entspricht. Umformungen, die in (1) Gleichungen miteinander kombinieren, werden in (2) also Zeilen der Matrix miteinander kombinieren. Wie wir jetzt sehen werden, lassen sich solche Zeilenumformungen sehr effizient als Matrixprodukte schreiben.

Konstruktion 14.1 (Elementarmatrizen)**.** Es seien $m, n \in \mathbb{N}$ und $A = (a_{i,j})_{i,j} \in K^{m \times n}$.

(a) Für $k \in \{1, \dots, m\}$ und $\lambda \in K \backslash \{0\}$ setzen wir

$$F_k(\lambda) := \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \lambda & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \in K^{m \times m},$$

wobei der Eintrag λ in Zeile und Spalte k steht. Die Matrix $F_k(\lambda)$ ist also nichts weiter als die Einheitsmatrix, bei der der Eintrag 1 in der k-ten Zeile und Spalte durch ein $\lambda \neq 0$ ersetzt wurde. Mit dieser Matrix ist

$$F_k(\lambda) \cdot A = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \lambda & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{k,1} & \cdots & a_{k,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ \lambda a_{k,1} & \cdots & \lambda a_{k,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix},$$

d. h. die Multiplikation von A mit $F_k(\lambda)$ von links entspricht genau der Multiplikation der k-ten Zeile von A mit λ.

(b) Für $k,l \in \{1,\dots,m\}$ mit $k \neq l$ und $\lambda \in K$ setzen wir

$$F_{k,l}(\lambda) := \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \lambda & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \in K^{m\times m},$$

wobei der Eintrag λ in Zeile k und Spalte l steht, d. h. diesmal haben wir in der Einheitsmatrix den Eintrag 0 in Zeile k und Spalte l durch λ ersetzt. (Beachte, dass der Eintrag λ für $k < l$ oberhalb und für $k > l$ unterhalb der Diagonale steht; wir haben in der Matrix oben der Einfachheit halber nur den Fall $k < l$ dargestellt.) In diesem Fall ist

$$F_{k,l}(\lambda)\cdot A = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \lambda & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \vdots & & \vdots \\ a_{k,1} & \cdots & a_{k,n} \\ \vdots & & \vdots \\ a_{l,1} & \cdots & a_{l,n} \\ \vdots & & \vdots \end{pmatrix} = \begin{pmatrix} \vdots & & \vdots \\ a_{k,1}+\lambda a_{l,1} & \cdots & a_{k,n}+\lambda a_{l,n} \\ \vdots & & \vdots \\ a_{l,1} & \cdots & a_{l,n} \\ \vdots & & \vdots \end{pmatrix},$$

d. h. die Multiplikation von A mit $F_{k,l}(\lambda)$ von links entspricht der Addition des λ-fachen von Zeile l zu Zeile k.

Die Matrizen $F_k(\lambda)$ für $\lambda \in K\backslash\{0\}$ sowie $F_{k,l}(\lambda)$ für $k \neq l$ und $\lambda \in K$ heißen **Elementarmatrizen**. Es gibt sie in jeder (quadratischen) Größe $m \times m$; zur Vereinfachung der Schreibweise deuten wir diese Größe in der Notation $F_k(\lambda)$ bzw. $F_{k,l}(\lambda)$ aber nicht an.

Man sagt, dass $F_k(\lambda)\cdot A$ und $F_{k,l}(\lambda)\cdot A$ aus A durch eine **elementare Zeilenumformung** entstehen.

Bemerkung 14.2.

(a) Die Elementarmatrizen sind invertierbar mit

$$(F_k(\lambda))^{-1} = F_k\left(\frac{1}{\lambda}\right) \quad \text{und} \quad (F_{k,l}(\lambda))^{-1} = F_{k,l}(-\lambda).$$

Dies folgt direkt aus Konstruktion 14.1: Wenn wir z. B. das Matrixprodukt

$$F_k\left(\frac{1}{\lambda}\right)\cdot F_k(\lambda) = F_k\left(\frac{1}{\lambda}\right)\cdot F_k(\lambda)\cdot E$$

bilden, multiplizieren wir die k-te Zeile in der Einheitsmatrix zuerst mit λ und dann mit $\frac{1}{\lambda}$, d. h. es kommt insgesamt wieder die Einheitsmatrix heraus – also ist $F_k(\frac{1}{\lambda})\cdot F_k(\lambda) = E$. Genauso ergibt sich auch $F_k(\lambda)\cdot F_k(\frac{1}{\lambda})$, und damit ist $F_k(\frac{1}{\lambda})$ das Inverse von $F_k(\lambda)$.

Analog zeigt man die Aussage über das Inverse von $F_{k,l}(\lambda)$.

(b) Das Vertauschen von zwei Zeilen $k,l \in \{1,\dots,m\}$ mit $k \neq l$ in einer Matrix $A \in K^{m\times n}$ lässt sich als Folge von elementaren Zeilenumformungen realisieren: Sind $a_1,\dots,a_m$ die Zeilen von A, so erhalten wir

$$A = \begin{pmatrix} \vdots \\ a_k \\ \vdots \\ a_l \\ \vdots \end{pmatrix} \xrightarrow{F_{k,l}(1)\cdot} \begin{pmatrix} \vdots \\ a_k+a_l \\ \vdots \\ a_l \\ \vdots \end{pmatrix} \xrightarrow{F_{l,k}(-1)\cdot} \begin{pmatrix} \vdots \\ a_k+a_l \\ \vdots \\ -a_k \\ \vdots \end{pmatrix} \xrightarrow{F_{k,l}(1)\cdot} \begin{pmatrix} \vdots \\ a_l \\ \vdots \\ -a_k \\ \vdots \end{pmatrix} \xrightarrow{F_l(-1)\cdot} \begin{pmatrix} \vdots \\ a_l \\ \vdots \\ a_k \\ \vdots \end{pmatrix}.$$

(c) Führen wir mit einer Matrix $A \in K^{m\times n}$ mehrere elementare Zeilenumformungen aus, so ist das Ergebnis nach Konstruktion 14.1 eine Matrix $F_1\cdot \cdots \cdot F_r\cdot A$ mit Elementarmatrizen $F_1,\dots,F_r$. Da diese Elementarmatrizen nach (a) invertierbar sind, ist das Produkt

$F := F_1 \cdot \dots \cdot F_r$ nach Lemma 13.16 ebenfalls wieder invertierbar. Das Ergebnis der Zeilenumformungen ist also die Matrix FA, d. h. wir haben A von links mit einer invertierbaren Matrix multipliziert. In der Tat werden wir in Lemma 14.11 noch sehen, dass *jede* Multiplikation mit einer invertierbaren Matrix von links als Abfolge von elementaren Zeilenumformungen geschrieben werden kann.

Mit derartigen Zeilenumformungen wollen wir jetzt eine Matrix bzw. ein lineares Gleichungssystem auf eine möglichst einfache Form bringen. Die folgende Form wird dabei herauskommen.

Definition 14.3 (Zeilenstufenform)**.** Es seien $m,n,r \in \mathbb{N}$ und $A = (a_{i,j})_{i,j} \in K^{m\times n}$.

(a) Wir sagen, dass A in **Zeilenstufenform** mit r **Stufen** ist, wenn

$$A = \begin{array}{c} \begin{array}{cccccc} & k_1 & k_2 & & k_r & \\ & \downarrow & \downarrow & & \downarrow & \end{array} \\ \begin{pmatrix} & 1 & & & & \\ & & 1 & & & * \\ & & & \ddots & & \\ & 0 & & & 1 & \\ & & & & & \end{pmatrix} \end{array}$$

ist, wobei das Symbol „$*$" für beliebige Einträge steht. Mit anderen Worten gibt es also Spalten $1 \le k_1 < \cdots < k_r \le n$ (die sogenannten **Stufenspalten**), so dass gilt:

- $a_{i,k_i} = 1$ für alle $i = 1,\dots,r$;
- rechts hiervon stehen beliebige Einträge (d. h. $a_{i,j}$ ist beliebig falls $i \le r$ und $j > k_i$);
- alle anderen Einträge sind 0.

(b) Sind zusätzlich noch außer den geforderten Einsen alle anderen Einträge in den Stufenspalten gleich Null, ist A also sogar von der Form

$$\begin{pmatrix} 1 & * \cdots * & 0 & * \cdots * & & 0 & * \cdots * \\ & & 1 & * \cdots * & & 0 & * \cdots * \\ & & & & \ddots & \vdots & \vdots \\ & 0 & & & & 1 & * \cdots * \\ & & & & & & \end{pmatrix}$$

(d. h. ist $a_{l,k_i} = 0$ für alle $l < i \le r$), so sagt man, dass A in **reduzierter Zeilenstufenform** ist.

(c) Ist $S \in \mathrm{GL}(m,K)$ eine invertierbare Matrix, so dass SA in (reduzierter) Zeilenstufenform ist, so heißt SA eine (reduzierte) Zeilenstufenform von A.

Der wesentliche Satz in diesem Abschnitt besagt nun, dass jede Matrix eine (reduzierte) Zeilenstufenform besitzt. Dabei sind beide Formen in der Praxis wichtig: Die reduzierte Zeilenstufenform wird sich in Satz 14.6 als eindeutig herausstellen und ist damit aus theoretischer Sicht natürlich deutlich schöner – andererseits werden wir aber auch sehen, dass die normale, nicht-reduzierte Zeilenstufenform für viele Zwecke ausreichend und mit weniger Rechenaufwand zu erzielen ist. In der Tat ist der Beweis des folgenden Satzes konstruktiv in dem Sinne, dass er auch ein Verfahren angibt, wie eine (reduzierte) Zeilenstufenform mit elementaren Zeilenumformungen erreicht werden kann. Dieser Algorithmus im Beweis ist mindestens genauso wichtig wie die eigentliche Aussage:

Satz 14.4 (**Gauß-Algorithmus**)**.** *Jede Matrix $A \in K^{m\times n}$ lässt sich mit elementaren Zeilenumformungen auf (reduzierte) Zeilenstufenform bringen.*

Insbesondere besitzt A im Sinne von Definition 14.3 (c) also eine (reduzierte) Zeilenstufenform.

Beweis. Wir beweisen den Satz mit Induktion über die Anzahl n der Spalten von A. Da für $n = 0$ nichts zu zeigen ist, müssen wir nur den Induktionsschritt durchführen. Es sei also $A \in K^{m\times(n+1)}$ beliebig vorgegeben. Wir wollen A mit elementaren Zeilenumformungen auf (reduzierte) Zeilenstufenform bringen und nehmen nach Induktionsvoraussetzung an, dass wir dies für alle Matrizen mit n Spalten bereits können. Wir unterscheiden dabei zwei Fälle:

Fall 1: Alle Einträge in der ersten Spalte von A sind 0, d. h. A hat die Form

$$A = \left(\begin{array}{c|c} 0 & \\ \vdots & A' \\ 0 & \end{array}\right)$$

für eine Matrix $A' \in K^{m\times n}$. Nach Induktionsvoraussetzung können wir A' dann durch elementare Zeilenumformungen auf (reduzierte) Zeilenstufenform bringen. Da diese Zeilenumformungen an der ersten Nullspalte aber nichts ändern, haben wir damit auch A auf (reduzierte) Zeilenstufenform gebracht.

Fall 2: Es sind nicht alle Einträge in der ersten Spalte von A gleich 0.

(a) Falls der Eintrag $a_{1,1}$ links oben in A gleich Null ist, vertauschen wir zwei Zeilen von A so, dass dieser Eintrag nicht mehr gleich 0 ist (nach Bemerkung 14.2 (b) ist dies durch elementare Zeilenumformungen machbar).

(b) Wir dividieren die erste Zeile durch $a_{1,1}$ und erhalten eine Matrix der Form

$$\left(\begin{array}{c|c} 1 & \\ a_{2,1} & \\ \vdots & * \\ a_{m,1} & \end{array}\right).$$

(c) Wir subtrahieren von jeder Zeile $k \in \{2,\dots,m\}$ das $a_{k,1}$-fache der ersten Zeile und bekommen dadurch

$$\left(\begin{array}{c|c} 1 & * \cdots * \\ \hline 0 & \\ \vdots & A' \\ 0 & \end{array}\right)$$

mit einer Matrix $A' \in K^{(m-1)\times n}$.

(d) Nach Induktionsvoraussetzung können wir jetzt A' durch elementare Zeilenumformungen auf Zeilenstufenform bringen. Damit hat aber auch die gesamte Matrix bereits Zeilenstufenform – wollen wir also nur diese normale, nicht-reduzierte Zeilenstufenform erreichen, so sind wir damit fertig. Andernfalls bringen wir A' gemäß der Induktionsvoraussetzung sogar auf reduzierte Zeilenstufenform und bekommen eine Matrix der Form

$$\left(\begin{array}{c|cccccc} 1 & *\cdots* & * & *\cdots* & & * & *\cdots* \\ \hline 0 & & 1 & *\cdots* & & 0 & *\cdots* \\ \vdots & & & & \ddots & & \\ \vdots & & 0 & & & & \\ \vdots & & & & & 1 & *\cdots* \\ 0 & & & & & & \end{array}\right).$$

Subtrahieren wir nun geeignete Vielfache der Zeilen $2,\dots,m$ von der ersten Zeile, so können wir damit dann noch die Einträge in den Stufenspalten der ersten Zeile zu Null machen und erhalten so auch die reduzierte Zeilenstufenform. □

Beispiel 14.5. Wir wollen die Matrix

$$A = \begin{pmatrix} -1 & 2 & 4 \\ -1 & 3 & 7 \\ 1 & 1 & 5 \end{pmatrix} \in \mathbb{R}^{3\times 3}$$

mit elementaren Zeilenumformungen auf Zeilenstufenform bringen. Dazu können wir nach dem Algorithmus im Beweis von Satz 14.4 wie folgt vorgehen (die Notation $Z3-3Z2 \to Z3$ bedeutet z. B., dass wir das Dreifache der zweiten Zeile von der dritten subtrahieren):

$$\begin{pmatrix} -1 & 2 & 4 \\ -1 & 3 & 7 \\ 1 & 1 & 5 \end{pmatrix} \xrightarrow{-Z1\to Z1} \begin{pmatrix} 1 & -2 & -4 \\ -1 & 3 & 7 \\ 1 & 1 & 5 \end{pmatrix} \xrightarrow{\substack{Z2+Z1\to Z2 \\ Z3-Z1\to Z3}} \begin{pmatrix} 1 & -2 & -4 \\ 0 & 1 & 3 \\ 0 & 3 & 9 \end{pmatrix}$$

$$\xrightarrow{Z3-3Z2\to Z3} \begin{pmatrix} 1 & -2 & -4 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix}.$$

Möchten wir eine reduzierte Zeilenstufenform, so addieren wir schließlich noch das Doppelte der zweiten Zeile zur ersten und erhalten

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix}.$$

Satz 14.6 (Eindeutigkeit der reduzierten Zeilenstufenform)**.** *Die reduzierte Zeilenstufenform einer Matrix $A \in K^{m\times n}$ ist eindeutig.*

Beweis. Wie in Satz 14.4 zeigen wir diese Aussage wieder mit Induktion über die Anzahl n der Spalten von A, wobei auch hier der Induktionsanfang für $n = 0$ trivial ist.

Für den Induktionsschritt sei $n \in \mathbb{N}$ so, dass die reduzierte Zeilenstufenform jeder Matrix in $K^{m\times n}$ eindeutig ist. Wir betrachten nun eine Matrix in $K^{m\times(n+1)}$, die wir als $(A\,|\,b)$ mit $A \in K^{m\times n}$ und $b \in K^m$ schreiben können. Es seien

$$S(A\,|\,b) = (SA\,|\,Sb) \quad \text{und} \quad \tilde{S}(A\,|\,b) = (\tilde{S}A\,|\,\tilde{S}b)$$

mit $S, \tilde{S} \in \mathrm{GL}(m,K)$ zwei reduzierte Zeilenstufenformen dieser Matrix. Wir müssen zeigen, dass sie übereinstimmen.

Da eine reduzierte Zeilenstufenform durch Streichen der letzten Spalte wieder in eine reduzierte Zeilenstufenform übergeht, sind SA und $\tilde{S}A$ damit reduzierte Zeilenstufenformen von $A \in K^{m\times n}$. Nach Induktionsvoraussetzung gilt also bereits $SA = \tilde{S}A$, d. h. wir müssen nur noch $Sb = \tilde{S}b$ zeigen.

Es sei dazu r die Anzahl der Stufen in der Zeilenstufenform $SA = \tilde{S}A$. Insbesondere enthält diese Matrix dann die Einheitsvektoren $e_1,\dots,e_r$ in den Stufenspalten $k_1,\dots,k_r$ wie in Definition 14.3; nach Beispiel 13.10 (d) gilt also

$$SAe_{k_i} = \tilde{S}Ae_{k_i} = e_i \quad\Rightarrow\quad S^{-1}e_i = \tilde{S}^{-1}e_i \qquad \text{für alle } i = 1,\dots,r. \tag{$*$}$$

Wir unterscheiden nun zwei Fälle:

(a) Hat die reduzierte Zeilenstufenform $(SA\,|\,Sb)$ ebenfalls nur r Stufen, so hat Sb die Form $\lambda_1 e_1 + \cdots + \lambda_r e_r$ für gewisse $\lambda_1,\dots,\lambda_r \in K$. Wir erhalten dann wie behauptet

$$\tilde{S}b = \tilde{S}S^{-1}Sb = \tilde{S}S^{-1}(\lambda_1 e_1 + \cdots \lambda_r e_r) \overset{(*)}{=} \tilde{S}\tilde{S}^{-1}(\lambda_1 e_1 + \cdots \lambda_r e_r) = \lambda_1 e_1 + \cdots \lambda_r e_r = Sb.$$

Analog folgt dies natürlich, wenn die reduzierte Zeilenstufenform $(\tilde{S}A\,|\,\tilde{S}b)$ aus r Stufen besteht.

(b) Haben beide reduzierten Zeilenstufenformen $(SA\,|\,Sb)$ und $(\tilde{S}A\,|\,\tilde{S}b)$ mehr als r Stufen, so muss die letzte Spalte dieser Matrizen eine Stufenspalte sein. Nach Definition 14.3 (b) ist dies nur möglich für $Sb = \tilde{S}b = e_{r+1}$. □

Bemerkung 14.7 (Die reduzierte Zeilenstufenform als Normalform)**.** Oft betrachtet man in der linearen Algebra Matrizen mit einer bestimmten Art erlaubter Umformungen (in unserem Fall Multiplikationen mit invertierbaren Matrizen von links) und versucht damit, eine eindeutig bestimmte und in gewissem Sinne möglichst einfache Form der Matrix zu erreichen. Eine solche Form bezeichnet man dann als *Normalform*.

In dieser Sprechweise ist die reduzierte Zeilenstufenform nach Satz 14.4 und Satz 14.6 also eine Normalform bezüglich der Multiplikation mit invertierbaren Matrizen von links. Wir werden in dieser Vorlesung noch einige weitere Normalformen bezüglich anderer Umformungen kennenlernen (siehe **??**).

14.B Der Rang von Matrizen

Da jede Matrix A nach Satz 14.4 und Satz 14.6 eine eindeutig bestimmte reduzierte Zeilenstufenform besitzt, ist insbesondere natürlich auch die Anzahl der Stufen in dieser Form eindeutig durch A festgelegt. Diese Zahl, die für die Untersuchung von A ganz besonders wichtig ist, wollen wir in diesem Abschnitt nun genauer betrachten.

Definition 14.8 (Rang einer Matrix)**.** Für eine Matrix $A \in K^{m\times n}$ heißt die Anzahl der Stufen in der reduzierten Zeilenstufenform von A der **Rang** von A. Wir schreiben ihn als $\operatorname{rk} A$.

Bemerkung 14.9. Es sei $A \in K^{m\times n}$.

(a) Offensichtlich ist $\operatorname{rk} A \le m$ und auch $\operatorname{rk} A \le n$, da nicht mehr Stufen in einer Matrix der Größe $m \times n$ Platz haben.

(b) Wenden wir den Gauß-Algorithmus an, um eine Matrix in Zeilenstufenform auf eine Matrix in reduzierter Zeilenstufenform zu bringen, so wird die Matrix im Beweis von Satz 14.4 nur noch in Schritt (d) umgeformt, der die Anzahl der Stufen aber nicht mehr verändert.

Wollen wir den Rang einer Matrix A berechnen, genügt uns also schon eine *beliebige* Zeilenstufenform von A: Ihre Anzahl Stufen ist ebenfalls gleich $\operatorname{rk} A$. Wir werden dies im Folgenden oft verwenden, ohne jedes Mal explizit darauf hinzuweisen.

Beispiel 14.10.

(a) Für alle $m,n \in \mathbb{N}$ sind sowohl die Nullmatrix $0 \in K^{m\times n}$ als auch die Einheitsmatrix $E_n \in K^{n\times n}$ bereits in (reduzierter) Zeilenstufenform mit 0 bzw. n Stufen. Also ist $\operatorname{rk} 0 = 0$ und $\operatorname{rk} E_n = n$.

(b) Für die Matrix
$$A = \begin{pmatrix} -1 & 2 & 4 \\ -1 & 3 & 7 \\ 1 & 1 & 5 \end{pmatrix} \quad \in \mathbb{R}^{3\times 3}$$
aus Beispiel 14.5 gilt $\operatorname{rk} A = 2$, da die dort berechnete Zeilenstufenform zwei Stufen hat.

(c) Hat $A \in K^{m\times n}$ nur r Zeilen, in denen ein Eintrag ungleich 0 vorkommt, so ist $\operatorname{rk} A \le r$: Wir können dann diese r Zeilen nach oben tauschen und den Gauß-Algorithmus nur mit diesen obersten r Zeilen durchführen, wodurch sich natürlich auch eine Zeilenstufenform mit höchstens r Stufen ergibt.

Mit Hilfe des Rangs einer Matrix können wir nun bereits das Problem aus Kapitel 13 lösen, wie man von einer quadratischen Matrix bestimmen kann, ob sie invertierbar ist, und in diesem Fall die inverse Matrix berechnen kann.

Lemma 14.11 (Äquivalente Kriterien für Invertierbarkeit)**.** *Für eine quadratische Matrix $A \in K^{n\times n}$ sind äquivalent:*

(a) $A \in \mathrm{GL}(n,K)$*, d. h. A ist invertierbar.*

(b) $\operatorname{rk} A = n$.

(c) *Die reduzierte Zeilenstufenform von A ist E_n.*

(d) *A ist ein Produkt von Elementarmatrizen.*

Beweis.

(a) $\Rightarrow$ (b): Mit dem Gauß-Algorithmus finden wir ein Produkt F von Elementarmatrizen, so dass FA in Zeilenstufenform ist. Aber FA ist auch invertierbar (da F und A es sind) und kann damit nach Beispiel 13.18 (c) keine Zeile haben, deren Einträge alle 0 sind. Also muss es n Stufen in der Zeilenstufenform FA geben, d. h. es ist $\operatorname{rk} A = n$.

(b) $\Rightarrow$ (c): Dies ist klar, da E_n die einzige $n \times n$-Matrix in reduzierter Zeilenstufenform mit n Stufen ist.

(c) $\Rightarrow$ (d): Nach Voraussetzung können wir A mit dem Gauß-Algorithmus auf die Einheitsmatrix bringen, d. h. es gibt Elementarmatrizen $F_1, \dots, F_k$ mit $F_1 \cdots F_k A = E_n$. Nach Bemerkung 14.2 (a) sind Elementarmatrizen aber invertierbar und ihre Inversen wieder Elementarmatrizen, und damit ist $A = F_k^{-1} \cdots F_1^{-1}$ ein Produkt von Elementarmatrizen.

(d) $\Rightarrow$ (a): Da Elementarmatrizen invertierbar sind, sind nach Lemma 13.16 auch Produkte von Elementarmatrizen wieder invertierbar. □

Algorithmus 14.12 (Inverse Matrix)**.** Gegeben sei eine quadratische Matrix $A \in K^{n \times n}$. Wir wollen überprüfen, ob A invertierbar ist, und in diesem Fall die inverse Matrix A^{-1} berechnen. Dazu starten wir mit der $n \times (2n)$-Matrix $(A \,|\, E_n)$, in der wir A und die Einheitsmatrix der gleichen Größe nebeneinander schreiben. Wir bringen dann die Matrix A mit dem Gauß-Algorithmus auf reduzierte Zeilenstufenform, machen die dafür benötigten Zeilenumformungen aber in allen $2n$ Spalten der Matrix. Ist $F \in \mathrm{GL}(n, K)$ das Produkt der Elementarmatrizen, das den durchgeführten Umformungen entspricht, so erhalten wir also die Matrix $(FA \,|\, FE_n) = (FA \,|\, F)$, wobei in der linken Hälfte die reduzierte Zeilenstufenform FA von A steht.

Nach Lemma 14.11 ist A genau dann invertierbar, wenn für diese reduzierte Zeilenstufenform die Einheitsmatrix herausgekommen ist, wenn also $FA = E_n$ ist. In diesem Fall ist aber offensichtlich $F = A^{-1}$ die gesuchte inverse Matrix – und diese steht nach den Umformungen genau in der rechten Hälfte unseres Diagramms.

Als konkretes Beispiel wollen wir die Matrix

$$A = \begin{pmatrix} 1 & 1 \\ 3 & 4 \end{pmatrix} \quad \in \mathbb{R}^{2 \times 2}$$

wählen. Wir wenden also wie gewohnt den Gauß-Algorithmus auf die Matrix A an, führen aber alle Umformungen mit der 2×4-Matrix $(A \,|\, E_2)$ durch:

$$\left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right) \overset{Z2-3Z1 \to Z2}{\longrightarrow} \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 1 & -3 & 1 \end{array}\right) \overset{Z1-Z2 \to Z1}{\longrightarrow} \left(\begin{array}{cc|cc} 1 & 0 & 4 & -1 \\ 0 & 1 & -3 & 1 \end{array}\right).$$

Da die herausgekommene reduzierte Zeilenstufenform von A (die linke Hälfte dieser Matrix) die Einheitsmatrix ist, ist A invertierbar. Die inverse Matrix steht nun in der rechten Hälfte des Diagramms: Es ist

$$A^{-1} = \begin{pmatrix} 4 & -1 \\ -3 & 1 \end{pmatrix}.$$

Bisher haben wir mit einer gegebenen Matrix $A \in K^{m \times n}$ nur *Zeilen*umformungen durchgeführt, weil wir damit – wie wir am Anfang von Abschnitt 14.A schon gesehen hatten und gleich noch genauer untersuchen werden – gerade die möglichen Umformungen eines linearen Gleichungssystems der Form $Ax = b$ mit $x \in K^n$ und $b \in K^m$ beschreiben können. Diese Umformungen konnten wir durch Multiplikation mit einer invertierbaren Matrix $S \in \mathrm{GL}(m, K)$ von links erreichen, nämlich mit einem Produkt von Elementarmatrizen.

Natürlich können wir mit A aber analog auch *Spalten*umformungen durchführen, was dann wie in Konstruktion 14.1 einer Multiplikation mit einem Produkt von Elementarmatrizen bzw. einer invertierbaren Matrix $T \in \mathrm{GL}(n, K)$ von rechts entspricht. Erlauben wir sowohl Zeilen- als auch Spaltenumformungen, würden wir vermutlich erwarten, dass wir A auf eine noch einfachere Form als eine reduzierte Zeilenstufenform bringen können, da wir ja mehr Möglichkeiten für die Umformungen haben. In der Tat wollen wir jetzt sehen, dass wir dann eine noch viel einfachere Normalform im Sinne von Bemerkung 14.7 erhalten können: Der Rang kann durch diese Umformungen zwar nicht geändert werden, aber für einen gegebenen Rang erhalten wir immer dieselbe Normalform.

Diese Umformung von Matrizen durch Multiplikation mit invertierbaren Matrizen von links und rechts hat übrigens eine besondere Bedeutung, die wir in **??** noch sehen werden. Sie hat daher einen speziellen Namen.

Definition 14.13 (Äquivalente Matrizen)**.** Es seien $m,n \in \mathbb{N}$. Zwei Matrizen $A,B \in K^{m\times n}$ heißen **äquivalent** zueinander, wenn es invertierbare Matrizen $S \in \mathrm{GL}(m,K)$ und $T \in \mathrm{GL}(n,K)$ gibt mit $B = SAT$.

Man prüft leicht nach, dass dies in der Tat eine Äquivalenzrelation auf $K^{m\times n}$ gemäß Definition 2.27 ist.

Lemma 14.14. *Für jede Matrix $A \in K^{m\times n}$ gilt:*

(a) *Für alle $S \in \mathrm{GL}(m,K)$ gilt $\mathrm{rk}(SA) = \mathrm{rk}\,A$.*

(b) *Für alle $T \in K^{n\times p}$ gilt $\mathrm{rk}(AT) \le \mathrm{rk}\,A$, mit Gleichheit falls $T \in \mathrm{GL}(n,K)$.*

Insbesondere haben äquivalente Matrizen also denselben Rang.

Beweis. Es sei $F \in \mathrm{GL}(m,K)$ ein Produkt von Elementarmatrizen, so dass FA in Zeilenstufenform mit $r := \mathrm{rk}\,A$ Stufen ist.

(a) Wegen $FS^{-1} \cdot SA = FA$ ist die Matrix FA auch eine Zeilenstufenform von SA, es ist also auch $\mathrm{rk}(SA) = r = \mathrm{rk}\,A$.

(b) Da die unteren $m-r$ Zeilen von FA nur Nullen enthalten, gilt dies nach Definition der Matrixmultiplikation auch für das Matrixprodukt $FA \cdot T$. Nach Beispiel 14.10 (c) ist damit $\mathrm{rk}(AT) \overset{(a)}{=} \mathrm{rk}(FAT) \le r = \mathrm{rk}\,A$.

Ist T sogar invertierbar, folgt daraus auch $\mathrm{rk}\,A = \mathrm{rk}(AT \cdot T^{-1}) \le \mathrm{rk}(AT)$, und damit die Gleichheit $\mathrm{rk}(AT) = \mathrm{rk}\,A$. □

Satz 14.15 (Normalform von Matrizen bezüglich Äquivalenz)**.** *Zu jeder Matrix $A \in K^{m\times n}$ gibt es ein $r \in \mathbb{N}$ und invertierbare Matrizen $S \in \mathrm{GL}(m,K)$ und $T \in \mathrm{GL}(n,K)$ mit*

$$SAT = \left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right) = \left(\begin{array}{ccc|c} 1 & & 0 & \\ & \ddots & & 0 \\ 0 & & 1 & \\ \hline & 0 & & 0 \end{array}\right) \in K^{m\times n}.$$

Dabei ist r eindeutig bestimmt als $r = \mathrm{rk}\,A$.

*Analog zu Bemerkung 14.7 sagt man, dass eine solche Matrix in **Normalform** bezüglich der Äquivalenz von Matrizen ist.*

Beweis. Nach Lemma 14.14 ist $r = \mathrm{rk}(SAT) = \mathrm{rk}\,A$; eine Normalform wie im Satz ist also höchstens für $r = \mathrm{rk}\,A$ möglich.

Andererseits können wir eine solche Normalform aber auch einfach durch Zeilen- und Spaltenumformungen erreichen: Dazu bringen wir dei Matrix A zunächst mit elementaren Zeilenumformungen auf ihre reduzierte Zeilenstufenform. Da bei dieser Zeilenstufenform genau die Einheitsvektoren $e_1,\dots,e_r$ in den Stufenspalten stehen, können wir nun durch Subtraktion geeigneter Vielfacher dieser Stufenspalten von den anderen Spalten alle übrigen Einträge der Matrix zu 0 machen. Durch Spaltenvertauschung können wir schließlich noch die Stufenspalten ganz nach links schieben und so die gewünschte Normalform erreichen. □

Beispiel 14.16. Da die Matrix

$$A = \begin{pmatrix} -1 & 2 & 4 \\ -1 & 3 & 7 \\ 1 & 1 & 5 \end{pmatrix} \quad \in \mathbb{R}^{3\times 3}$$

aus Beispiel 14.5 den Rang 2 hat, gibt es nach Satz 14.15 invertierbare Matrizen $S,T \in \mathrm{GL}(3,\mathbb{R})$ mit

$$SAT = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Eine wichtige Folgerung aus Satz 14.15 ist, dass sich der Rang einer Matrix durch Transponieren nicht ändert: Für Matrizen in einer solchen Normalform ist dies nämlich offensichtlich, und jede andere ist ja äquivalent dazu. Exakt aufgeschrieben sieht dieser Beweis dann wie folgt aus.

Folgerung 14.17. *Für jede Matrix $A \in K^{m\times n}$ gilt $\mathrm{rk}(A^\mathsf{T}) = \mathrm{rk}\,A$.*

Beweis. Nach Satz 14.15 gibt es $S \in \mathrm{GL}(m,K)$ und $T \in \mathrm{GL}(n,K)$, so dass

$$SAT = \left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right)$$

mit $r = \mathrm{rk}\,A$. Mit Lemma 13.11 (d) gilt dann auch

$$T^\mathsf{T} A^\mathsf{T} S^\mathsf{T} = (SAT)^\mathsf{T} = \left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right)^\mathsf{T} = \left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right)$$

(wobei sich beim letzten Gleichheitszeichen die Größe der Matrix von $m \times n$ auf $n \times m$ ändert). Nach Folgerung 13.17 (d) sind mit S und T ferner auch S^T und T^T invertierbar, d. h. $T^\mathsf{T} A^\mathsf{T} S^\mathsf{T}$ ist äquivalent zu A^T. Damit ergibt sich aus Lemma 14.14

$$\mathrm{rk}\,A^\mathsf{T} = \mathrm{rk}(T^\mathsf{T} A^\mathsf{T} S^\mathsf{T}) = \mathrm{rk}\left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right) = r = \mathrm{rk}\,A. \qquad \square$$

Eine weitere Folgerung hieraus ist die in Beispiel 13.18 (a) schon angekündigte Aussage, dass es für die Bestimmung der Invertierbarkeit einer quadratischen Matrix ausreicht, eine links- oder rechtsinverse Matrix zu finden.

Folgerung 14.18 (Rang eines Matrixprodukts)**.** *Es seien $A \in K^{m\times n}$ und $B \in K^{n\times p}$.*

(a) *Es gilt $\mathrm{rk}(AB) \le \mathrm{rk}\,A$ und $\mathrm{rk}(AB) \le \mathrm{rk}\,B$.*

(b) *Für quadratische Matrizen (also $m = n = p$) genügt bereits eine der beiden Gleichungen $AB = E_n$ und $BA = E_n$ dafür, dass A invertierbar ist mit $A^{-1} = B$.*

Beweis.

(a) Die erste Ungleichung ist genau Lemma 14.14 (b). Die zweite erhalten wir daraus nun einfach mit Folgerung 14.17 durch Transponieren, denn

$$\mathrm{rk}(AB) = \mathrm{rk}\left((AB)^\mathsf{T}\right) = \mathrm{rk}(B^\mathsf{T} A^\mathsf{T}) \overset{14.14\,(b)}{\le} \mathrm{rk}(B^\mathsf{T}) = \mathrm{rk}\,B.$$

(b) Es sei zunächst $AB = E_n$. Nach (a) ist dann $n = \mathrm{rk}\,E_n = \mathrm{rk}(AB) \le \mathrm{rk}\,A$, also $\mathrm{rk}\,A = n$. Nach Lemma 14.11 ist A damit invertierbar, und Multiplikation der Gleichung $AB = E_n$ mit A^{-1} von links liefert sofort $B = A^{-1}$.

Den Fall $BA = E_n$ zeigt man natürlich analog mit der zweiten Ungleichung aus (a). $\square$

14.C Lösungsmengen linearer Gleichungssysteme

Nach unseren Vorarbeiten in diesem Kapitel können wir jetzt untersuchen, wie die Lösungsmengen linearer Gleichungssysteme aussehen und konkret berechnet werden können. Dazu führen wir zunächst ein paar nützliche Notationen ein.

Definition 14.19 ($L(A,b)$, Bild und Kern einer Matrix)**.** Es seien $A \in K^{m\times n}$ eine Matrix und $b \in K^m$.

(a) Wir bezeichnen die Lösungsmenge des linearen Gleichungssystems $Ax = b$ mit

$$L(A,b) := \{x \in K^n : Ax = b\} \quad \subset K^n.$$

(b) Das **Bild** von A ist definiert als

$$\operatorname{Im} A := \{Ax : x \in K^n\} \quad \subset K^m$$

(die Schreibweise kommt vom englischen Wort „image“). Nach Definition ist das Gleichungssystem $Ax = b$ also genau dann lösbar, d. h. $L(A,b) \neq \emptyset$, wenn $b \in \operatorname{Im} A$.

(c) Der **Kern** von A ist definiert als

$$\operatorname{Ker} A := \{x \in K^n : Ax = 0\} \quad \subset K^n$$

(die Schreibweise kommt vom englischen Wort „kernel“). Offensichtlich ist $\operatorname{Ker} A = L(A,0)$.

Bemerkung 14.20. Schreiben wir in Definition 14.19 die Matrix A spaltenweise als $A = (a_1 \,|\, \cdots \,|\, a_n)$ mit $a_1, \dots, a_n \in K^m$, sowie den Vektor $x \in K^n$ als $x = (\lambda_j)_j$, so ist das dort auftretende Matrixprodukt Ax nach Definition gleich

$$Ax = (a_1 \,|\, \cdots \,|\, a_n) \cdot \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \lambda_1 a_1 + \cdots + \lambda_n a_n.$$

Formulieren wir die Konzepte vom Bild und Kern einer Matrix mit Hilfe dieser Spaltenvektoren $a_1, \dots, a_n$, so erhalten wir die folgenden sehr wichtigen Begriffe.

Definition 14.21 (Linearkombinationen und lineare Abhängigkeit)**.** Es seien $a_1, \dots, a_n \in K^m$ gegebene Vektoren und $A := (a_1 \,|\, \cdots \,|\, a_n) \in K^{m\times n}$.

(a) Für $\lambda_1, \dots, \lambda_n \in K$ heißt der Vektor

$$\lambda_1 a_1 + \cdots + \lambda_n a_n \quad \in K^m$$

eine **Linearkombination** von $a_1, \dots, a_n$. Die Menge aller dieser Linearkombinationen bezeichnen wir mit $\operatorname{Lin}(a_1, \dots, a_n)$; nach Definition 14.19 (b) und Bemerkung 14.20 ist offensichtlich $\operatorname{Lin}(a_1, \dots, a_n) = \operatorname{Im} A$.

(b) Die Vektoren $a_1, \dots, a_n$ heißen **linear abhängig**, wenn man aus ihnen eine Linearkombination des Nullvektors

$$\lambda_1 a_1 + \cdots + \lambda_n a_n = 0 \qquad (*)$$

bilden kann, in der mindestens ein $\lambda_1, \dots, \lambda_n$ ungleich 0 ist (man nennt dies auch eine *nichttriviale Linearkombination* des Nullvektors).

Ist das Gegenteil der Fall, folgt aus der Linearkombination $(*)$ des Nullvektors also bereits, dass $\lambda_1 = \cdots = \lambda_n = 0$ gelten muss, so heißen die Vektoren $a_1, \dots, a_n$ **linear unabhängig**. In der Matrixschreibweise von Definition 14.19 (c) und Bemerkung 14.20 bedeutet dies offensichtlich genau $\operatorname{Ker} A = \{0\}$.

Beispiel 14.22.

(a) Die Einheitsvektoren $e_1,\dots,e_n \in K^n$ sind linear unabhängig: Sind nämlich $\lambda_1,\dots,\lambda_n \in K$ mit

$$\lambda_1 e_1 + \cdots + \lambda_n e_n = 0, \quad \text{also} \quad \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

so folgt daraus natürlich sofort $\lambda_1 = \cdots = \lambda_n = 0$.

(b) Enthalten die Vektoren $a_1,\dots,a_n \in K^m$ den Nullvektor, also ist $a_i = 0$ für ein $i \in \{1,\dots,n\}$, so sind diese Vektoren immer linear abhängig, denn dann ist ja $1 \cdot a_i = 0$ eine nicht-triviale Linearkombination des Nullvektors.

Ebenso sind die Vektoren $a_1,\dots,a_n$ immer linear abhängig, wenn ein Vektor unter ihnen mehrfach vorkommt, also wenn $a_i = a_j$ für gewisse $i,j \in \{1,\dots,n\}$ mit $i \neq j$ gilt: Dann ist nämlich $1 \cdot a_i - 1 \cdot a_j = 0$ eine nicht-triviale Linearkombination des Nullvektors.

(c) Es seien $a_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $a_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, $a_3 = \begin{pmatrix} 3 \\ 1 \end{pmatrix} \in \mathbb{R}^2$.

Dann sind a_1 und a_2 linear unabhängig, denn aus

$$\lambda_1 a_1 + \lambda_2 a_2 = 0 \quad \Rightarrow \quad \lambda_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Rightarrow \quad \lambda_1 + 2\lambda_2 = \lambda_1 + \lambda_2 = 0$$

folgt sofort $\lambda_1 = \lambda_2 = 0$. Genauso sieht man, dass a_1 und a_3 linear unabhängig sind, und ebenso a_2 und a_3.

Aber alle drei Vektoren a_1, a_2, a_3 zusammen sind linear abhängig, denn es gilt

$$1 \cdot a_1 - 2 \cdot a_2 + 1 \cdot a_3 = 1 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 2 \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} + 1 \cdot \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Die lineare Unabhängigkeit von Vektoren $a_1,\dots,a_n$ ist also eine Eigenschaft *aller dieser Vektoren zusammen* – wir können sie nicht überprüfen, indem wir uns die Vektoren nacheinander einzeln oder in Paaren anschauen.

Bemerkung 14.23 (Untervektorräume und Dimension)**.** Die Mengen $\mathrm{Lin}(a_1,\dots,a_n)$ aus Definition 14.21 (a) bestehen nach Definition aus allen skalaren Vielfachen und Summen der gegebenen Vektoren $a_1,\dots,a_n$. Gemäß Beispiel 13.4 (b) können wir sie uns damit vorstellen als Geraden, Ebenen oder entsprechende „höherdimensionale Mengen" durch den Nullpunkt. Solche Mengen werden wir in **??** als Untervektorräume kennenlernen und noch genau untersuchen.

Sind die Vektoren $a_1,\dots,a_n$ dabei linear unabhängig, so bedeutet dies anschaulich, dass sie „in unabhängige Richtungen zeigen" und damit „einen n-dimensionalen Raum aufspannen" – auch diesen Dimensionsbegriff werden wir in **??** noch exakt einführen. Für den Moment wollen wir nur festhalten, dass jeder Vektor in $\mathrm{Lin}(a_1,\dots,a_n)$ dann *auf eindeutige Art* eine Linearkombination der gegebenen Vektoren $a_1,\dots,a_n$ ist, so dass wir uns $\mathrm{Lin}(a_1,\dots,a_n)$ also als eine „Menge mit n unabhängigen Parametern" vorstellen können:

Lemma 14.24 (Eindeutigkeit von Linearkombinationen)**.** *Es seien $a_1,\dots,a_n \in K^m$ linear unabhängige Vektoren. Dann gibt es zu jedem Vektor $v \in \mathrm{Lin}(a_1,\dots,a_n)$ eindeutig bestimmte $\lambda_1,\dots,\lambda_n \in K$ mit $v = \lambda_1 a_1 + \cdots + \lambda_n a_n$.*

Beweis. Es seien $\lambda_1,\dots,\lambda_n \in K$ sowie $\tilde\lambda_1,\dots,\tilde\lambda_n \in K$ mit

$$v = \lambda_1 a_1 + \cdots + \lambda_n a_n = \tilde\lambda_1 a_1 + \cdots + \tilde\lambda_n a_n.$$

Daraus folgt natürlich sofort

$$(\lambda_1 - \tilde\lambda_1) a_1 + \cdots + (\lambda_n - \tilde\lambda_n) a_n = 0,$$

wegen der linearen Unabhängigkeit der Vektoren also $\lambda_i - \tilde\lambda_i = 0$ und damit auch $\lambda_i = \tilde\lambda_i$ für alle $i \in \{1,\dots,n\}$. □

Wir kommen nun wie angekündigt zur Hauptaussage über die Lösungsmengen linearer Gleichungssysteme. Genau wie beim Gauß-Algorithmus ist auch der Beweis dieses Satzes konstruktiv – er gibt also auch an, wie die dort mit $v_1, \dots, v_{n-r}$ und v bezeichneten Vektoren konkret berechnet werden können.

Satz 14.25 (Lösungsmengen linearer Gleichungssysteme). *Es seien $A \in K^{m\times n}$ eine Matrix vom Rang $r := \operatorname{rk} A$ und $b \in K^m$.*

(a) *Es gibt $n-r$ linear unabhängige Vektoren $v_1, \dots, v_{n-r} \in K^n$ mit*
$$L(A,0) = \operatorname{Ker} A = \operatorname{Lin}(v_1, \dots, v_{n-r}).$$

(b) *Das Gleichungssystem $Ax = b$ hat genau dann eine Lösung $v \in L(A,b)$, wenn die sogenannte* erweiterte Koeffizientenmatrix *$(A\,|\,b) \in K^{m\times(n+1)}$ ebenfalls den Rang r hat.*

(c) *Ist $v \in L(A,b)$ eine Lösung des Gleichungssystems $Ax = b$, so ist die gesamte Lösungsmenge gegeben durch $L(A,b) = v + \operatorname{Ker} A$.*

Insbesondere ist das lineare Gleichungssystem $Ax = b$ also genau dann eindeutig lösbar, wenn $\operatorname{rk}(A\,|\,b) = \operatorname{rk} A = n$ gilt.

Beweis. Wir bringen die erweiterte Koeffizientenmatrix $(A\,|\,b)$ mit dem Gauß-Algorithmus auf reduzierte Zeilenstufenform $S(A\,|\,b) = (SA\,|\,Sb)$ mit $S \in \operatorname{GL}(m,K)$. Dies ändert weder den Rang der betrachteten Matrizen (siehe Lemma 14.14 (a)) noch die gesuchten Lösungsmengen (wegen $Ax = b \Leftrightarrow SAx = Sb$). Das neue Gleichungssystem hat nun die erweiterte Koeffizientenmatrix

$$
\begin{array}{c}
\begin{array}{ccccccc} k_1 & & k_2 & & & k_r & \\ \downarrow & & \downarrow & & & \downarrow & \end{array}\\
\left(\begin{array}{ccccccc|c}
1 & * \cdots * & 0 & * \cdots * & & 0 & * \cdots * & b_1\\
 & & 1 & * \cdots * & & 0 & * \cdots * & b_2\\
 & & & & \ddots & \vdots & \vdots & \vdots\\
 & \mathbf{0} & & & & 1 & * \cdots * & b_r\\
 & & & & & & & \vdots\\
 & & & & & & & b_m
\end{array}\right)
\end{array}, \tag{1}
$$

wobei $(b_i)_i := Sb$ ist und „$*$“ die Einträge von $(a_{i,j})_{i,j} := SA$ bezeichnet, die ungleich 0 sein können. Wir bezeichnen die Stufenspalten von SA mit $k_1 \le \dots \le k_r$, und die Nichtstufenspalten von SA mit $j_1 \le \dots \le j_{n-r}$.

(a) Ist $b = 0$, so ist die i-te Zeile des umgeformten Gleichungssystems für $i \in \{1, \dots, r\}$
$$x_{k_i} + \sum_{l=1}^{n-r} a_{i,j_l} x_{j_l} = 0 \quad \Leftrightarrow \quad x_{k_i} = -\sum_{l=1}^{n-r} a_{i,j_l} x_{j_l}. \tag{2}$$
Die Gleichungen bestimmen also die Stufenvariablen $x_{k_1}, \dots, x_{k_r}$ eindeutig aus den Nichtstufenvariablen $x_{j_1}, \dots, x_{j_{n-r}}$. Wir können daher $\lambda_l := -x_{j_l} \in K$ für $l = 1, \dots, n-r$ als freie Parameter wählen, und erhalten damit aus (2) als Vektor geschrieben die allgemeine Lösung
$$x = \lambda_1 v_1 + \dots + \lambda_{n-r} v_{n-r}, \tag{3}$$
wobei v_l der Vektor ist mit j_l-ter Koordinate -1 und k_i-ter Koordinate a_{i,j_l} für $i = 1, \dots, r$ (alle anderen Koordinaten sind 0).

Damit ist also $L(A,0) = \operatorname{Lin}(v_1, \dots, v_{n-r})$. Außerdem ergibt sich in (3) für x nur dann der Nullvektor, wenn alle $\lambda_l = -x_{j_l}$ gleich 0 sind. Damit sind $v_1, \dots, v_{n-r}$ auch linear unabhängig.

(b) Die Matrix $(SA\,|\,Sb)$ in reduzierter Zeilenstufenform in (1) kann nur Rang r oder $r+1$ haben.

Hat sie Rang $r+1$, so ist $b_{r+1} = 1$, und die $(r+1)$-te Zeile des umgeformten Gleichungssystems liefert den Widerspruch $0 = 1$. In diesem Fall ist also $L(A,b) = \emptyset$.

Hat sie Rang r, so erhalten wir offensichtlich eine Lösung, indem wir die Nichtstufenvariablen $x_{j_1},\dots,x_{j_{n-r}}$ gleich 0 setzen, und die Stufenvariablen gleich $x_{k_i} = b_i$ für $i \in \{1,\dots,r\}$.

(c) Ist $Av = b$, so ist für ein $x \in K^n$ genau dann $Ax = b$, wenn $A(x-v) = 0$ gilt, also $x - v \in \operatorname{Ker} A$ und damit $x \in v + \operatorname{Ker} A$ ist. □

Bemerkung 14.26.

(a) Die konkrete Form der Vektoren $v_1,\dots,v_{n-r}$ und v in Satz 14.25 lässt sich gut zu einer Merkregel zusammenfassen: Nach der Berechnung der reduzierten Zeilenstufenform $(SA\,|\,Sb) \in K^{m\times(n+1)}$ bilden wir eine neue Matrix $C \in K^{n\times(n+1)}$, bei der wir für alle $i \in \{1,\dots,r\}$ die Zeile i von $(SA\,|\,Sb)$ in die Zeile k_i von C schreiben. Die übrigen Einträge von C setzen wir gleich -1 auf der Diagonalen und 0 sonst.

Nach dem Beweis von (a) stehen dann in den Spalten $j_1,\dots,j_{n-r}$ von C, die den Nichtstufenspalten von SA entsprechen, die Vektoren $v_1,\dots,v_{n-r}$: In Spalte j_l steht in Zeile k_i der Eintrag a_{i,j_l}, in Zeile j_l der Eintrag -1, und sonst überall 0.

In der letzten Spalte steht nach dem Beweis von (b) außerdem der Vektor v.

(b) Falls wir nur ein Gleichungssystem $Ax = 0$ mit rechter Seite 0 lösen (also $\operatorname{Ker} A$ bestimmen) wollen, müssen wir in (a) die rechte Spalte für b gar nicht erst hinschreiben, da die Einträge dort ohnehin alle gleich 0 sind und während des Verfahrens auch bleiben. In diesem Fall ist das Gleichungssystem natürlich auch in jedem Fall lösbar mit $x = 0$.

(c) Alternativ zum Verfahren im Beweis des Satzes kann man die erweiterte Koeffizientenmatrix $(A\,|\,b)$ auch nur auf nicht notwendig reduzierte Zeilenstufenform bringen (was natürlich weniger Rechenarbeit ist). Im Fall der Lösbarkeit bestimmt dann die i-te Gleichung des umgeformten Gleichungssystems für $i \in \{1,\dots,r\}$ die Stufenvariable x_{k_i} aus den Nichtstufenvariablen *und den weiter rechts stehenden Stufenvariablen* $x_{k_{i+1}},\dots,x_{k_r}$. Wir müssen diese Gleichungen dann also von unten nach oben durchgehen und so der Reihe nach $x_{k_r},\dots,x_{k_1}$ aus den freien Variablen *und den bereits bestimmten Stufenvariablen* berechnen (was dann wieder etwas mehr Arbeit ist).

Beispiel 14.27. Wir wollen $\operatorname{Ker} A = L(A,0)$ und $L(A,b)$ für

$$A = \begin{pmatrix} 1 & 4 & 0 & 5 \\ 2 & 8 & 1 & 7 \\ 0 & 0 & 2 & -6 \end{pmatrix} \in \mathbb{R}^{3\times 4} \quad \text{und} \quad b = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix} \in \mathbb{R}^3$$

bestimmen. Dazu bringen wir $(A\,|\,b)$ auf reduzierte Zeilenstufenform:

$$\left(\begin{array}{cccc|c} 1 & 4 & 0 & 5 & 1 \\ 2 & 8 & 1 & 7 & 3 \\ 0 & 0 & 2 & -6 & 2 \end{array}\right) \xrightarrow{Z2-2Z1\to Z2} \left(\begin{array}{cccc|c} 1 & 4 & 0 & 5 & 1 \\ 0 & 0 & 1 & -3 & 1 \\ 0 & 0 & 2 & -6 & 2 \end{array}\right) \xrightarrow{Z3-2Z2\to Z3} \left(\begin{array}{cccc|c} 1 & 4 & 0 & 5 & 1 \\ 0 & 0 & 1 & -3 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Das gegebene Gleichungssystem ist also äquivalent zu

$$\begin{aligned} x_1 + 4x_2 \quad + 5x_4 &= 1, \\ x_3 - 3x_4 &= 1, \\ 0 &= 0. \end{aligned}$$

Offensichtlich lässt sich dieses umgeformte Gleichungssystem nun leicht lösen: Die Variablen x_2 und x_4 können wir frei wählen; die erste Gleichung bestimmt daraus dann eindeutig x_1 und die zweite x_3.

Um die so entstehende Lösung mit dem Verfahren aus Bemerkung 14.26 (a) direkt hinschreiben zu können, schreiben wir die beiden Zeilen der umgeformten Matrix wie unten grau hinterlegt in die Zeilen 1 und 3 einer 4×5-Matrix C (da die Stufen in den Spalten 1 und 3 liegen), und füllen die restlichen Einträge mit -1 auf der Diagonalen und 0 sonst. Da die Nichtstufenspalten der obigen Zeilenstufenform die zweite und vierte sind, lesen wir dann die Vektoren v_1, v_2 und v in Satz 14.25 direkt in der Matrix C in den eingekreisten Spalten 2 und 4 bzw. der letzten Spalte ab: Mit

$$C = \left(\begin{array}{ccccc|c} 1 & 4 & 0 & 5 & 1 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -3 & 1 \\ 0 & 0 & 0 & -1 & 0 \end{array}\right) \quad \Rightarrow \quad v_1 = \begin{pmatrix} 4 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \; v_2 = \begin{pmatrix} 5 \\ 0 \\ -3 \\ -1 \end{pmatrix}, \; v = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

ist $\operatorname{Ker} A = \operatorname{Lin}(v_1, v_2)$ (mit linear unabhängigen Vektoren v_1 und v_2) und $L(A,b) = v + \operatorname{Lin}(v_1, v_2)$.

Zum Abschluss dieses Kapitels wollen wir noch einige einfache, aber sehr wichtige Folgerungen aus unseren Ergebnissen festhalten. Die erste betrifft die sogenannte universelle Lösbarkeit linearer Gleichungssysteme – d. h. die Frage nach der Lösbarkeit eines solchen Gleichungssystems $Ax = b$ für eine beliebige rechte Seite b.

Folgerung 14.28 (Universelle Lösbarkeit linearer Gleichungssysteme)**.** *Für jede Matrix $A \in K^{m \times n}$ gilt:*

(a) $\operatorname{rk} A = n \;\Leftrightarrow\;$ *Für alle $b \in K^m$ hat das Gleichungssystem $Ax = b$ höchstens eine Lösung.*

(b) $\operatorname{rk} A = m \;\Leftrightarrow\;$ *Für alle $b \in K^m$ hat das Gleichungssystem $Ax = b$ mindestens eine Lösung.*

Beweis.

(a) Nach Satz 14.25 (a) ist $\operatorname{rk} A = n$ genau dann, wenn $\operatorname{Ker} A = \{0\}$ gilt, also wenn das Gleichungssystem $Ax = 0$ genau die Lösung $x = 0$ hat. Nach Satz 14.25 (c) hat dann auch jedes Gleichungssystem $Ax = b$ höchstens eine Lösung.

(b) „$\Rightarrow$“: Ist $\operatorname{rk} A = m$, so ist wegen $(A\,|\,b) \in K^{m \times (n+1)}$ auch $\operatorname{rk}(A\,|\,b) = m$ nach Bemerkung 14.9 (a). Damit ist $L(A,b) \neq \emptyset$ nach Satz 14.25 (b).

„$\Leftarrow$“: Ist $\operatorname{rk} A < m$, so hat die reduzierte Zeilenstufenform SA von A unten (mindestens) eine Nullzeile. Damit hat das Gleichungssystem $SAx = e_m \Leftrightarrow Ax = S^{-1} e_m$ dann keine Lösung. □

Bemerkung 14.29. Nach Folgerung 14.28 hat das Gleichungssystem $Ax = b$ mit $A \in K^{m \times n}$ und $b \in K^m$ genau dann für alle b eine eindeutige Lösung, wenn $\operatorname{rk} A = m = n$, also wenn A quadratisch und invertierbar ist (siehe Lemma 14.11). Diesen einfachen Fall hatten wir schon in Folgerung 13.17 (d) betrachtet: Die eindeutige Lösung ist dann $x = A^{-1} b$.

Unsere Ergebnisse zeigen auch, dass die lineare Unabhängigkeit von Vektoren $a_1, \ldots, a_n \in K^m$ eng mit dem Rang der daraus gebildeten Matrix $(a_1\,|\cdots|\,a_n) \in K^{m \times n}$ zusammenhängt. In der Tat liefert Folgerung 14.31 zusammen mit der geometrischen Interpretation der linearen Unabhängigkeit aus Bemerkung 14.23 vermutlich eine der anschaulichsten Arten, sich den Rang einer Matrix (ohne Bezug auf algorithmische Konzepte wie den Gauß-Algorithmus oder eine Zeilenstufenform) anschaulich vorzustellen.

Folgerung 14.30 (Rangkriterium für lineare Unabhängigkeit)**.** *Es seien $a_1, \ldots, a_n \in K^m$.*

(a) *Die Vektoren $a_1, \ldots, a_n$ sind genau dann linear unabhängig, wenn $\operatorname{rk}(a_1\,|\cdots|\,a_n) = n$ gilt.*

(b) *Für $n > m$ sind die Vektoren $a_1, \ldots, a_n$ in jedem Fall linear abhängig.*

Beweis. Es sei $A = (a_1\,|\cdots|\,a_n) \in K^{m \times n}$ mit Rang $r := \operatorname{rk} A$.

(a) Die Vektoren $a_1, \ldots, a_n$ sind nach Definition 14.21 (b) genau dann linear unabhängig, wenn $\operatorname{Ker} A = \{0\}$. Dies ist nach Satz 14.25 (a) äquivalent zu $r = n$.

(b) Aus $n > m$ folgt mit Bemerkung 14.9 (a) sofort $\operatorname{rk} A \le m < n$. Nach (a) sind $a_1, \ldots, a_n$ dann also linear abhängig. □

Folgerung 14.31 (Rang = Maximalzahl linear unabhängiger Spalten)**.** *Für jede Matrix $A \in K^{m \times n}$ ist $\operatorname{rk} A$ die Maximalzahl linear unabhängiger Spalten in A.*

Beweis. Für beliebige Spalten $1 \le k_1 \le \cdots \le k_s \le n$ sei $A' = (a_{k_1}\,|\cdots|\,a_{k_s}) \in K^{m \times s}$ die Teilmatrix von A, die durch Auswahl dieser s Spalten entsteht.

Wir bringen A nun auf Zeilenstufenform $SA = (Sa_1 \,|\, \cdots \,|\, Sa_n)$ mit $r := \operatorname{rk} A$ Stufen, und betrachten die Matrix
$$SA' = (Sa_{k_1} \,|\, \cdots \,|\, Sa_{k_s}) \quad \in K^{m \times s},$$
die ebenfalls durch Auswahl der entsprechenden s Spalten in SA entsteht:

- Ist $s = r$ und wählen wir für $k_1, \dots, k_r$ genau die Stufenspalten, so ist SA' ebenfalls in Zeilenstufenform mit r Stufen. Also ist dann $\operatorname{rk} A' = r = s$, und damit sind die s ausgewählten Spalten von A nach Folgerung 14.30 (a) linear unabhängig.
- Ist $s > r$, so hat SA' genau wie SA höchstens r Zeilen mit Einträgen ungleich 0, und damit folgt $\operatorname{rk} A' \leq r < s$ aus Beispiel 14.10 (c). Also sind die s ausgewählten Spalten nach Folgerung 14.30 (a) diesmal linear abhängig.

Insgesamt ist r damit die Maximalzahl linear unabhängiger Spalten in A. □

Bemerkung 14.32. Der Beweis von Folgerung 14.31 ist in dem Sinne konstruktiv, dass er auch angibt, wie man r linear unabhängige Spalten aus einer Matrix A vom Rang r auswählen kann: Man bringt A auf Zeilenstufenform und wählt dann die Spalten der Ausgangsmatrix A, die zu den Stufenspalten gehören. So haben wir z. B. für die Matrix
$$A = \begin{pmatrix} -1 & 2 & 4 \\ -1 & 3 & 7 \\ 1 & 1 & 5 \end{pmatrix} = (a_1 \,|\, a_2 \,|\, a_3) \in \mathbb{R}^{3\times 3} \quad \text{mit} \quad a_1 = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix},\; a_2 = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix},\; a_3 = \begin{pmatrix} 4 \\ 7 \\ 5 \end{pmatrix}$$
in Beispiel 14.5 eine Zeilenstufenform mit 2 Stufen und Stufenspalten 1 und 2 (und damit $\operatorname{rk} A = 2$) bestimmt. Dementsprechend sind die Vektoren a_1 und a_2 also nach Folgerung 14.31 linear unabhängig, während alle drei Vektoren a_1, a_2, a_3 linear abhängig sind.

Wir haben in Folgerung 14.30 (b) auch gesehen, dass in K^m bis zu maximal m Vektoren linear unabhängig sein können. Wenn wir also $n < m$ linear unabhängige Vektoren $a_1, \dots, a_n$ in K^m haben (so wie hier im Beispiel a_1 und a_2 in K^3), können wir uns daher fragen, ob wir diese Vektoren immer zu m linear unabhängigen Vektoren ergänzen können. Wie wir jetzt sehen werden, ist dies in der Tat immer möglich – sogar mit Einheitsvektoren, und wiederum mit einem konstruktiven Verfahren.

Lemma 14.33 (Ergänzung linear unabhängiger Vektoren)**.** *Es seien $a_1, \dots, a_n \in K^m$ linear unabhängige Vektoren. Dann gibt es Einheitsvektoren $e_{j_1}, \dots, e_{j_{m-n}} \in K^m$, so dass die m Vektoren $a_1, \dots, a_n, e_{j_1}, \dots, e_{j_{m-n}}$ linear unabhängig sind.*

Beweis. Wir setzen wieder $A = (a_1 \,|\, \cdots \,|\, a_n) \in K^{m \times n}$; nach Folgerung 14.30 (a) ist also $\operatorname{rk} A = n$. Diesmal bringen wir A mit elementaren *Spalten*umformungen auf eine Spaltenstufenform mit n Stufen. Sind nun $j_1, \dots, j_{m-n}$ die Nichtstufenzeilen dieser Spaltenstufenform, so können wir aus A durch Hinzufügen von Spalten mit den Einheitsvektoren $e_{j_1}, \dots, e_{j_{m-n}}$ eine $m \times m$-Matrix in Spaltenstufenform mit m Stufen erzeugen. Da diese Matrix dann Rang m hat, sind die m Vektoren $a_1, \dots, a_n, e_{j_1}, \dots, e_{j_{m-n}}$ damit nach Folgerung 14.30 (a) linear unabhängig. □

Beispiel 14.34. Wir hatten in Bemerkung 14.32 gesehen, dass die dort angegebenen Vektoren $a_1, a_2 \in \mathbb{R}^3$ linear unabhängig sind. Wollen wir a_1 und a_2 durch Hinzunahme eines weiteren Vektors zu drei linear unabhängigen Vektoren ergänzen, bringen wir die Matrix $A = (a_1 \,|\, a_2)$ wie in Lemma 14.33 auf Spaltenstufenform mit 2 Stufen:
$$A = \begin{pmatrix} -1 & 2 \\ -1 & 3 \\ 1 & 1 \end{pmatrix} \xrightarrow{-S1 \to S1} \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ -1 & 1 \end{pmatrix} \xrightarrow{S2 - 2S1 \to S2} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ -1 & 3 \end{pmatrix}.$$
Da die dritte Zeile hierbei keine Stufenzeile ist, sind die Vektoren a_1, a_2, e_3 linear unabhängig.

Beachte, dass dies nicht die einzige Möglichkeit wäre: In diesem Beispiel könnte man leicht nachrechnen, dass auch a_1, a_2, e_1 und a_1, a_2, e_2 linear unabhängig sind. Der Algorithmus aus Lemma 14.33 liefert nur eine mögliche Art, die gegebenen linear unabhängigen Vektoren mit Einheitsvektoren zu maximal vielen linear unabhängigen Vektoren zu ergänzen.

15. Vektorräume

In den letzten beiden Kapiteln haben wir ausführlich die Lösungsmengen linearer Gleichungssysteme untersucht – und dabei sowohl theoretische Ergebnisse über ihre Struktur bewiesen als auch Verfahren für ihre Berechnung entwickelt. Dafür hat es sich als praktisch erwiesen, die Sprache der Vektoren in K^m bzw. Matrizen in $K^{m\times n}$ (mit $m,n \in \mathbb{N}$) zu benutzen.

Wie schon in Bemerkung 13.6 erwähnt ist der Begriff des Vektors in der Mathematik jedoch viel allgemeiner. Analog zu Gruppen und Körpern in Kapitel 3.A werden Vektoren nämlich eigentlich über die mit ihnen möglichen Rechenoperationen definiert: Während es in einer Gruppe *eine* Verknüpfung und in einem Körper *zwei* Verknüpfungen „$+$“ und „$\cdot$“ mit den erwarteten Eigenschaften gibt (siehe Definition 3.1 und 3.5), sind es für Vektoren eine Addition und eine Skalarmultiplikation mit Elementen eines fest gewählten Grundkörpers K. Die bisher betrachteten Vektoren in K^m (und auch Matrizen) erlauben diese Rechenoperationen natürlich wie in Definition 13.3, aber viele andere Objekte auch: So können wir z. B. auch reelle Funktionen (punktweise) addieren und mit einer reellen Zahl multiplizieren. In diesem Sinne können wir also auch solche Funktionen als Vektoren bezeichnen, bzw. sie als Elemente eines sogenannten Vektorraums auffassen.

Der Vorteil dieses viel allgemeineren Vektorbegriffs liegt auf der Hand: Auf diese Art können wir die Ergebnisse, die wir über Vektoren (also nur aus der Existenz einer Vektoraddition und Skalarmultiplikation) herleiten können, in viel mehr Fällen anwenden. In der Tat ist die Untersuchung allgemeiner Vektorräume, mit der wir jetzt beginnen wollen, das eigentliche Ziel der linearen Algebra. Unsere Resultate der letzten beiden Kapitel werden uns helfen, zu den dabei eingeführten theoretischen Konzepten auch gleich viele Beispiele betrachten und konkret berechnen zu können.

15.A Der Vektorraumbegriff

Wie oben schon erwähnt wollen wir Vektoren über die Existenz einer Addition und Skalarmultiplikation definieren. Diese beiden Rechenoperationen sollen dabei natürlich gewisse erwartete Eigenschaften erfüllen – und zwar genau die, die wir für Vektoren in K^m und Matrizen in $K^{m\times n}$ in Lemma 13.5 bereits gesehen hatten:

Definition 15.1 (Vektorräume)**.** Es sei K ein Körper. Ein **Vektorraum** über K (oder K-Vektorraum) ist eine Menge V zusammen mit zwei Verknüpfungen

$$\begin{aligned} &+ : V \times V \to V \quad (\textit{Vektoraddition}) \\ \text{und} \qquad &\cdot : K \times V \to V \quad (\textit{Skalarmultiplikation}), \end{aligned}$$

so dass gilt:

(a) $(V,+)$ ist eine abelsche Gruppe (siehe Definition 3.1).

(b) (1. Distributivität) Für alle $\lambda,\mu \in K$ und $x \in V$ gilt $(\lambda+\mu)\cdot x = \lambda\cdot x + \mu\cdot x$.

(c) (2. Distributivität) Für alle $\lambda \in K$ und $x,y \in V$ gilt $\lambda\cdot(x+y) = \lambda\cdot x + \lambda\cdot y$.

(d) (Assoziativität) Für alle $\lambda,\mu \in K$ und $x \in V$ gilt $(\lambda\cdot\mu)\cdot x = \lambda\cdot(\mu\cdot x)$.

(e) Für alle $x \in V$ gilt $1\cdot x = x$.

Die Elemente von V heißen **Vektoren**, die Elemente von K **Skalare**.

Bemerkung 15.2.

(a) Beachte, dass man einen Vektorraum nur dann definieren kann, wenn man vorher einen Körper K gewählt hat. Wenn klar ist, welcher Körper gemeint ist, werden wir jedoch auch oft nur von einem Vektorraum (statt einem K-Vektorraum) sprechen.

(b) Wie wir es auch in den letzten beiden Kapiteln schon getan haben, haben wir in Definition 15.1 mehrfach die gleichen Symbole für unterschiedliche Dinge verwendet: Es gibt z. B. zwei Additionen, die wir beide mit „+“ bezeichnet haben, nämlich die Addition $+ \colon K \times K \to K$ zweier Körperelemente und die Addition $+ \colon V \times V \to V$ der Vektoren. Da man aus der Art der verknüpften Elemente eindeutig ablesen kann, um welche Verknüpfung es sich handeln muss, können dadurch aber keine Mehrdeutigkeiten entstehen: So werden z. B. beim ersten Pluszeichen in Definition 15.1 (b) zwei Skalare, beim zweiten jedoch zwei Vektoren addiert. Nur wenn wir auch in der Notation explizit deutlich machen wollen, um welche der beiden Verknüpfungen es sich handelt, schreiben wir diese als $+_K$ bzw. $+_V$. Analog gibt es auch die Multiplikation zweimal, einmal als Multiplikation $\cdot_K$ in K und einmal als Skalarmultiplikation $\cdot_V$, und auch zweimal die Null, nämlich einmal als Null 0_K im Körper K und einmal als *Nullvektor* 0_V, d. h. als das neutrale Element von $(V,+)$. In dieser ausführlichen Notation könnte man z. B. die Bedingung aus Definition 15.1 (b) als

$$(\lambda +_K \mu) \cdot_V x = \lambda \cdot_V x +_V \mu \cdot_V x$$

schreiben. In der Regel werden wir diese Indizes K und V jedoch weglassen, genauso wie die Malzeichen sowohl für $\cdot_K$ als auch für $\cdot_V$.

Beispiel 15.3.

(a) Für jeden Körper K ist $V = \{0\}$ (mit den trivialen Verknüpfungen) ein K-Vektorraum, der sogenannte **Nullvektorraum**.

(b) Für alle $m,n \in \mathbb{N}$ ist der Raum $K^{m\times n}$ aller $m \times n$-Matrizen über K nach Lemma 13.5 ein K-Vektorraum. Insbesondere ist damit auch der Raum K^m, dessen Elemente wir bisher als Vektoren bezeichnet haben, ein Vektorraum. Unser neuer Vektorbegriff ist also eine Verallgemeinerung der vorläufigen Notation 13.2.

Auch mit dem neuen Vektorbegriff bleiben die Vektorräume K^m für $m \in \mathbb{N}$ sicher die wichtigsten Beispiele für K-Vektorräume; wir werden sie also auch in Zukunft weiterhin oft als Beispiele betrachten. Im Fall $m = 1$ erhält man daraus $K^1 = K$, also K selbst als K-Vektorraum; der Fall $m = 0$ wird konventionsgemäß als der Nullvektorraum $K^0 = \{0\}$ aufgefasst.

(c) Sind V und W zwei K-Vektorräume, so ist auch ihr Produkt $V \times W$ mit komponentenweiser Addition und Skalarmultiplikation ein K-Vektorraum – der Nachweis der Vektorraumeigenschaften ist analog zum bereits bekannten Spezialfall $K^2 = K \times K$ (siehe Lemma 13.5). Genauso sind natürlich auch Produkte von mehr als zwei Vektorräumen wieder ein Vektorraum.

(d) Es seien K ein Körper und M eine Menge. Dann ist die Menge $\mathrm{Abb}(M,K) := \{f \colon M \to K\}$ aller Abbildungen von M nach K ein K-Vektorraum, indem wir Addition und Multiplikation punktweise definieren als

$$(f+g)(x) := f(x)+g(x) \quad \text{und} \quad (\lambda f)(x) := \lambda f(x)$$

für alle $\lambda \in K$, $x \in M$ und $f,g \colon M \to K$. In der Tat haben wir in Aufgabe 3.12 (a) bereits gesehen, dass $\mathrm{Abb}(M,K)$ eine abelsche Gruppe ist (der Nullvektor ist hierbei die Funktion, die jedes Element von M auf 0 abbildet, und das zu einer Funktion $f \colon M \to K$ additive Inverse die Funktion $-f \colon M \to K$, $x \mapsto -f(x)$). Die anderen Vektorraumeigenschaften zeigt man wieder analog.

Ein Spezialfall hiervon ist der Raum $\mathrm{Abb}(\mathbb{N},K)$, dessen Elemente $f \in \mathrm{Abb}(\mathbb{N},K)$ wir in Verallgemeinerung von (b) als „unendliche Folgen“ $(f(0), f(1), f(2), \dots)$ mit Elementen in K schreiben können. Im Fall $K = \mathbb{R}$ sind dies gerade die in der Analysis betrachteten reellen Zahlenfolgen.

Darüber hinaus lässt sich auf die gleiche Art auch die Menge $\mathrm{Abb}(M,W)$ aller Abbildungen von einer beliebigen Menge M in einen K-Vektorraum W zu einem Vektorraum machen.

(e) Die reellen Zahlen $\mathbb{R}$ bilden einen $\mathbb{Q}$-Vektorraum. In der Tat kann man reelle Zahlen addieren und mit einer rationalen multiplizieren, und es ist klar, dass mit diesen Definitionen alle Vektorraumeigenschaften gelten.

Analog zum Fall von Gruppen und Körpern wollen wir auch hier zunächst ein paar elementare Eigenschaften von Vektorräumen zeigen. Sie haben einen ähnlichen Charakter wie die Axiome in Definition 15.1, folgen aber bereits aus diesen (so dass man sie nicht separat fordern muss).

Lemma 15.4 (Eigenschaften von Vektorräumen)**.** *In jedem K-Vektorraum V gilt für alle $\lambda \in K$ und $x \in V$:*

(a) $0_K \cdot x = \lambda \cdot 0_V = 0_V$.

(b) *Ist $\lambda \cdot x = 0_V$, so ist $\lambda = 0_K$ oder $x = 0_V$.*

(c) $(-1) \cdot x = -x$.

Beweis.

(a) Der Beweis ist ganz analog zu dem von Lemma 3.8 (a): Wegen der 1. Distributivität aus Definition 15.1 (b) gilt $0_K \cdot x = (0_K + 0_K) \cdot x = 0_K \cdot x + 0_K \cdot x$, nach Subtraktion von $0_K \cdot x$ also wie behauptet $0_V = 0_K \cdot x$. Analog zeigt man $0_V = \lambda \cdot 0_V$ mit Hilfe der 2. Distributivität.

(b) Ist $\lambda x = 0_V$ und $\lambda \neq 0_K$, so folgt

$$\begin{aligned} x &= 1 \cdot x && \text{(Definition 15.1 (e))} \\ &= (\lambda^{-1} \cdot \lambda) x \\ &= \lambda^{-1}(\lambda x) && \text{(Definition 15.1 (d))} \\ &= 0_V && \text{(Teil (a)).} \end{aligned}$$

(c) Es gilt

$$\begin{aligned} (-1) \cdot x + x &= (-1) \cdot x + 1 \cdot x && \text{(Definition 15.1 (e))} \\ &= (-1+1) \cdot x && \text{(Definition 15.1 (b))} \\ &= 0_K \cdot x = 0_V && \text{(Teil (a)),} \end{aligned}$$

also ist $(-1) \cdot x$ das additive Inverse zu x. □

15.B Untervektorräume

Immer wenn man eine neue mathematische Struktur (wie z. B. Gruppen, Körper, oder jetzt hier die Vektorräume) einführt, sollte man als Erstes zwei Dinge untersuchen:

- die sogenannten *Unterstrukturen*, d. h. Teilmengen, die selbst wieder die betrachtete Struktur haben; und
- die sogenannten *Morphismen*, d. h. Abbildungen, die diese Struktur erhalten.

Wir haben dies in Kapitel **??** nur deswegen für Gruppen und Körper nicht getan, weil wir in dieser Vorlesung nur Vektorräume, aber nicht Gruppen und Körper ausführlich studieren wollen. Diejenigen von euch, die auch die Vorlesung „Algebraische Strukturen" hören, haben dort aber sicher schon z. B. Untergruppen und Morphismen von Gruppen untersucht – und werden jetzt feststellen, dass sich Untervektorräume und Morphismen von Vektorräumen, die wir nun studieren wollen, in vielen Aspekten sehr ähnlich verhalten.

Wir beginnen dabei mit den Untervektorräumen, durch die wir gleichzeitig auch sehr viele neue Beispiele von Vektorräumen erhalten. Morphismen werden wir danach in Abschnitt **??** untersuchen.

Definition 15.5 (Untervektorräume)**.** Es sei V ein Vektorraum über einem Körper K. Eine Teilmenge $U \subset V$ heißt **Untervektorraum** oder **Unterraum** von V, in Zeichen $U \leq V$, wenn „U mit den Verknüpfungen in V selbst wieder ein Vektorraum ist", d.h. wenn gilt:

(a) Für alle $x, y \in U$ und $\lambda \in K$ gilt auch $x + y \in U$ und $\lambda x \in U$ (d. h. die Addition und Skalarmultiplikation in V lassen sich auf U einschränken).

(b) U ist mit diesen eingeschränkten Verknüpfungen selbst ein K-Vektorraum.

Man sagt für (a) auch, dass U bezüglich der Vektoraddition und Skalarmultiplikation *abgeschlossen* ist. Beachte, dass diese Bedingung notwendig ist, um (b) überhaupt formulieren zu können, weil wir sonst ja gar keine Verknüpfungen auf U hätten.

Da das Nachprüfen der vielen Vektorraumaxiome in Definition 15.1 in der Praxis natürlich sehr aufwändig ist, sieht es vielleicht so aus, als ob das Nachprüfen der Unterraumeigenschaft (b) genauso aufwändig ist. Glücklicherweise ist dies jedoch nicht so: Wir wollen jetzt zeigen, dass sich praktisch alle Vektorraumaxiome von V direkt auf U übertragen, so dass man fast nur die Abgeschlossenheit in (a) nachprüfen muss.

Satz 15.6 (**Unterraumkriterium**). *Eine Teilmenge U eines K-Vektorraums V ist genau dann ein Unterraum von V, wenn gilt:*

(a) *Für alle $x,y \in U$ und $\lambda \in K$ gilt auch $x+y \in U$ und $\lambda x \in U$.*

(b) *$U \neq \emptyset$.*

Beweis. Da die Abgeschlossenheit (a) in Definition 15.5 und Satz 15.6 dieselbe Bedingung ist, müssen wir nur unter dieser Bedingung die Äquivalenz der Eigenschaften (b) überprüfen.

„$\Rightarrow$“: Ist U ein Vektorraum, so enthält U natürlich den Nullvektor und ist damit nicht leer.

„$\Leftarrow$“: Es sei nun $U \neq \emptyset$. Wir müssen die Vektorraumaxiome für U überprüfen.

- Assoziativität der Vektoraddition: Weil V ein Vektorraum ist, gilt $(x+y)+z = x+(y+z)$ für alle $x,y,z \in V$ und damit erst recht für alle $x,y,z \in U$. Die Assoziativität der Addition überträgt sich also direkt von V auf U.
- Additives neutrales Element: Nach Voraussetzung gibt es ein Element $x \in U$. Damit ist nach der Abgeschlossenheit der Skalarmultiplikation auch $0 = 0 \cdot x \in U$, und für diesen Nullvektor gilt natürlich $x+0 = 0+x = x$ für alle $x \in U$ (denn dies gilt ja sogar für alle $x \in V$).
- Additive inverse Elemente: Für jedes $x \in U$ ist wegen der Abgeschlossenheit der Skalarmultiplikation auch $(-1) \cdot x \in U$, und dies ist nach Lemma 15.4 (c) genau das additive inverse Element zu x.

Also ist $(U,+)$ schon einmal eine Gruppe. Die übrigen Axiome (also die Kommutativität der Vektoraddition und die Bedingungen (b) bis (e) aus Definition 15.1) sind aber alle von der Form, dass für alle Vektoren aus U eine bestimmte Gleichung gelten muss – und dies folgt nun genauso wie die Assoziativität oben sofort daraus, dass diese Gleichungen sogar für alle Vektoren aus V gelten. □

Bemerkung 15.7. Wie wir im Beweis von Satz 15.6 schon gesehen haben, muss jeder Unterraum U eines Vektorraums V den Nullvektor enthalten: Wegen $U \neq \emptyset$ gibt es ein $x \in U$, und aus der Abgeschlossenheit der Skalarmultiplikation folgt dann auch $0 = 0 \cdot x \in U$.

Die Bedingung $U \neq \emptyset$ in Satz 15.6 (b) ist also äquivalent zu $0 \in U$.

Beispiel 15.8.

(a) Für jeden Vektorraum V sind der Nullvektorraum $\{0\} \subset V$ und der gesamte Raum $V \subset V$ natürlich stets Unterräume von V. Sie werden die **trivialen Unterräume** von V genannt.

(b) Für alle $m,n \in \mathbb{N}$ und gegebene Vektoren $x_1,\dots,x_n \in K^m$ ist die Menge $\mathrm{Lin}(x_1,\dots,x_n)$ aller ihrer Linearkombinationen ein Unterraum von K^m: Sind nämlich

$$x = \lambda_1 x_1 + \cdots + \lambda_n x_n \quad \text{und} \quad y = \tilde{\lambda}_1 x_1 + \cdots + \tilde{\lambda}_n x_n$$

solche Linearkombinationen und $\lambda \in K$, so sind auch

$$x+y = (\lambda_1 + \tilde{\lambda}_1)x_1 + \cdots + (\lambda_n + \tilde{\lambda}_n)x_n \quad \text{und} \quad \lambda x = (\lambda\lambda_1)x_1 + \cdots + (\lambda\lambda_n)x_n$$

in $\mathrm{Lin}(x_1,\dots,x_n)$. Außerdem ist diese Menge nicht leer, da sie natürlich (konventionsgemäß auch für $n = 0$) den Nullvektor enthält. Damit ist $\mathrm{Lin}(x_1,\dots,x_n)$ nach Satz 15.6 ein Unterraum von K^m. Man nennt ihn den von $x_1,\dots,x_n$ **erzeugten** oder **aufgespannten Unterraum** von K^m.

Insbesondere gilt damit $\operatorname{Im} A \le K^m$ und $\operatorname{Ker} A \le K^n$ für jede Matrix $A \in K^{m\times n}$, da Bild und Kern einer Matrix nach Definition 14.21 (a) bzw. Satz 14.25 (a) eine solche Darstellung als Menge aller Linearkombinationen bestimmter Vektoren besitzen.

(c) Für $A \in K^{m\times n}$ und $b \in K^m$ betrachten wir die Lösungsmenge $L(A,b) \subset K^n$ des linearen Gleichungssystems $Ax = b$:

- Für $b = 0$ ist $L(A,0) = \operatorname{Ker} A$ nach (b) ein Unterraum.
- Für $b \neq 0$ dagegen ist $L(A,b)$ nach Bemerkung 15.7 niemals ein Unterraum, denn dann gilt ja $0 \notin L(A,b)$ wegen $A \cdot 0 = 0 \neq b$.

  In der Tat ist die Lösungsmenge – sofern sie nicht leer ist – dann nach Satz 14.25 (c) von der Form $v + \operatorname{Ker} A$ für ein $v \in K^n$ und damit ein *verschobener Unterraum*. Manchmal wird eine solche Menge in der Literatur auch als *affiner Unterraum* bezeichnet.

Als konkretes Beispiel können wir eine Matrix $A = (a_1 \;\; a_2) \in \mathbb{R}^{1\times 2}$ vom Rang 1 (d. h. $A \neq 0$) wählen. Dann ist wie im Bild unten der Unterraum

$$U_1 := \operatorname{Ker} A = \{x \in \mathbb{R}^2 : a_1 x_1 + a_2 x_2 = 0\}$$

(wobei x_1 und x_2 die Koordinaten von x bezeichnen) eine Ursprungsgerade, während für ein $b \in \mathbb{R}\backslash\{0\}$ die Gerade

$$U_2 := L(A,b) = \{x \in \mathbb{R}^2 : a_1 x_1 + a_2 x_2 = b\}$$

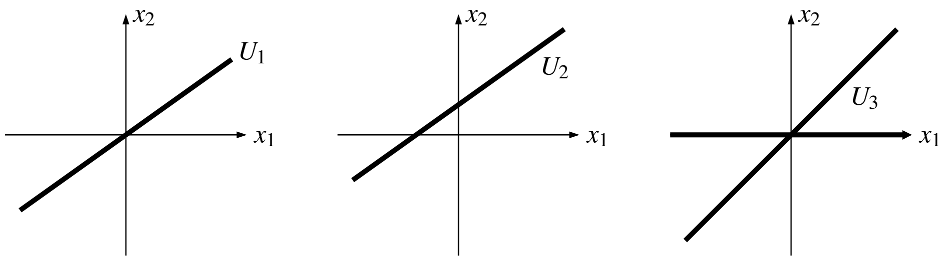
nicht durch den Ursprung läuft und damit kein Unterraum ist. Auch eine Vereinigung von zwei Ursprungsgeraden wie z. B.

$$U_3 := \{x \in \mathbb{R}^2 : x_2 = 0 \text{ oder } x_1 - x_2 = 0\}$$

ist kein Unterraum, denn hier ist z. B. wegen

$$\begin{pmatrix}1\\0\end{pmatrix} \in U_3 \quad \text{und} \quad \begin{pmatrix}1\\1\end{pmatrix} \in U_3, \quad \text{aber} \quad \begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}2\\1\end{pmatrix} \notin U_3$$

die Abgeschlossenheit der Vektoraddition nicht erfüllt.



$$\left\{\begin{pmatrix}x_1\\x_2\end{pmatrix} : x_2 = m x_1\right\} \qquad \left\{\begin{pmatrix}x_1\\x_2\end{pmatrix} : x_2 = m x_1 + 1\right\} \qquad \left\{\begin{pmatrix}x_1\\x_2\end{pmatrix} : x_2 = x_1 \text{ oder } x_2 = 0\right\}$$

Beachte, dass man hier nicht nur an den Rechnungen, sondern auch an den Bildern oben schon sehen kann, ob die gegebenen Teilmengen abgeschlossen bezüglich Vektoraddition und Skalarmultiplikation, also ob sie Untervektorräume sind.

(d) Eine quadratische Matrix $A \in K^{n\times n}$ heißt *symmetrisch*, wenn $A^{\mathsf{T}} = A$, also $a_{i,j} = a_{j,i}$ für alle $i,j \in \{1,\dots,n\}$ gilt. Mit dieser Definition ist die Teilmenge

$$U := \{A \in K^{n\times n} : A \text{ ist symmetrisch}\}$$

aller symmetrischen Matrizen ein Unterraum von $K^{n\times n}$: Es ist $0 \in U$, und für alle $A,B \in U$ (also $A^{\mathsf{T}} = A$ und $B^{\mathsf{T}} = B$) und $\lambda \in K$ gilt nach Lemma 13.7

$$(A+B)^{\mathsf{T}} = A^{\mathsf{T}} + B^{\mathsf{T}} = A + B \quad \text{und} \quad (\lambda A)^{\mathsf{T}} = \lambda (A^{\mathsf{T}}) = \lambda A,$$

also $A + B \in U$ und $\lambda A \in U$.

(e) Für eine gegebene Teilmenge $D \subset \mathbb{R}$ ist die Teilmenge $U \subset \mathrm{Abb}(D,\mathbb{R})$ aller Polynomfunktionen $f: D \to \mathbb{R}$ ein Unterraum des Vektorraums $\mathrm{Abb}(D,\mathbb{R})$ aus Beispiel 15.3 (d), denn mit f und g sind auch $f+g$ und λf für alle $\lambda \in \mathbb{R}$ Polynomfunktionen. Genauso ist für festes $k \in \mathbb{N}$ auch die Menge aller Polynomfunktionen vom Grad höchstens k ein Unterraum von $\mathrm{Abb}(D,\mathbb{R})$.

Im Fall des Vektorraums K^m können wir in der Tat sogar zeigen, dass *jeder* Unterraum die Form wie in Beispiel 15.3 (b) hat, also von endlich vielen Vektoren erzeugt werden kann – und zwar sogar von linear unabhängigen:

Satz 15.9 (Unterräume von K^n)**.** *Jeder Unterraum $U \le K^m$ kann als $U = \mathrm{Lin}(x_1,\dots,x_n)$ für linear unabhängige Vektoren $x_1,\dots,x_n \in K^m$ mit $n \in \mathbb{N}$ geschrieben werden.*

Beweis. Angenommen, es gäbe einen Unterraum $U \le K^m$, der nicht auf diese Art geschrieben werden kann. Wir könnten dann wie folgt rekursiv unendlich viele linear unabhängige Vektoren $x_1,x_2,x_3,\dots$ in U konstruieren:

Sind linear unabhängige $x_1,\dots,x_n \in U$ für ein $n \in \mathbb{N}$ bereits konstruiert, so gilt wegen der Abgeschlossenheit von U zunächst einmal auch $\mathrm{Lin}(x_1,\dots,x_n) \subset U$. Nach unserer Annahme kann aber nicht $U = \mathrm{Lin}(x_1,\dots,x_n)$ gelten, also gibt es ein $x_{n+1} \in U \backslash \mathrm{Lin}(x_1,\dots,x_n)$. Dann sind aber auch $x_1,\dots,x_{n+1}$ linear unabhängig: Sind nämlich $\lambda_1,\dots,\lambda_{n+1} \in K$ mit

$$\lambda_1 x_1 + \cdots \lambda_{n+1} x_{n+1} = 0 \quad \Rightarrow \quad \lambda_{n+1} x_{n+1} = -\lambda_1 x_1 - \cdots - \lambda_n x_n \in \mathrm{Lin}(x_1,\dots,x_n),$$

so folgt wegen $x_{n+1} \notin \mathrm{Lin}(x_1,\dots,x_n)$ zunächst $\lambda_{n+1} = 0$, und wegen der linearen Unabhängigkeit von $x_1,\dots,x_n$ dann auch $\lambda_1 = \cdots = \lambda_n = 0$.

Auf diese Art könnten wir also beliebig viele linear unabhängige Vektoren in U finden – was natürlich ein Widerspruch ist, da in K^m nach Folgerung 14.30 (b) höchstens m Vektoren linear unabhängig sein können. □

Bemerkung 15.10 (Berechnung von Unterräumen von K^m)**.** Wenn wir im Folgenden sagen, dass wir einen Unterraum von K^m *berechnen* wollen, meinen wir damit, ihn so darzustellen wie in Satz 15.9 – also linear unabhängige Vektoren zu bestimmen, die ihn erzeugen (beachte dabei, dass solche Erzeuger natürlich nicht eindeutig bestimmt sind, denn es ist ja z. B. $\mathrm{Lin}(e_1) = \mathrm{Lin}(2e_1)$).

Für alle in Beispiel 15.8 (b) erwähnten Unterräume können wir dies leicht durchführen:

(a) Ist der Unterraum gegeben als $\mathrm{Lin}(x_1,\dots,x_n)$ mit nicht notwendig linear unabhängigen $x_1,\dots,x_n$, so können wir aus diesen Vektoren mit Folgerung 14.31 **??** linear unabhängige Erzeuger auswählen.

(b) Ist der Unterraum gegeben als Bild einer Matrix $A = (x_1 | \cdots | x_n) \in K^{m\times n}$, so können wir wegen $\mathrm{Im}\,A = \mathrm{Lin}(x_1,\dots,x_n)$ genauso vorgehen.

(c) Ist der Unterraum gegeben als Kern einer Matrix, so liefert Satz 14.25 (a) sofort die gewünschte Darstellung.

Darüber hinaus haben wir in Aufgabe **??** gesehen, dass wir auch umgekehrt einen Unterraum der Form $\mathrm{Lin}(x_1,\dots,x_n)$ von K^m – nach Satz 15.9 also jeden Unterraum – immer als Kern einer Matrix A schreiben können. Dies liefert eine alternative Art der Darstellung beliebiger Unterräume von K^m, nämlich durch definierende Gleichungen $\{x \in K^m : Ax = 0\}$ statt durch erzeugende Vektoren $\{x = \lambda_1 x_1 + \cdots + \lambda_n x_n : \lambda_1,\dots,\lambda_n \in K\}$. Für konkrete Rechnungen mit Unterräumen kann jeweils die eine oder andere Darstellung besser geeignet sein.

Wir wollen nun sehen, wie man aus mehreren Unterräumen eines Vektorraums neue Unterräume konstruieren kann. Eine Möglichkeit besteht dabei einfach darin, ihren Durchschnitt zu bilden. Im Gegensatz dazu ist ihre Vereinigung nach Beispiel 15.8 (c) zwar in der Regel kein Unterraum; es gibt aber trotzdem eine Möglichkeit, aus ihnen einen neuen zu erzeugen, der sie enthält – die korrekte Konstruktion hierfür ist nur nicht die Vereinigung, sondern die sogenannte Summe von Unterräumen.

Lemma 15.11 (Durchschnitt und Summe von Unterräumen)**.** *Es seien U_1 und U_2 Unterräume eines Vektorraums V. Dann gilt:*

(a) *Der Durchschnitt $U_1 \cap U_2$ ist ebenfalls ein Unterraum von V.*

(b) *Die* ***Summe*** *$U_1 + U_2 := \{x_1 + x_2 : x_1 \in U_1, x_2 \in U_2\}$ ist ebenfalls ein Unterraum von V.*

Analog gilt dies auch für Durchschnitte $U_1 \cap \cdots \cap U_n$ und Summen $U_1 + \cdots + U_n$ von mehr als zwei Unterräumen.

Beweis. Wir überprüfen jeweils das Unterraumkriterium aus Satz 15.6. Beachte dabei zunächst, dass der Nullvektor nach Bemerkung 15.7 sowohl in U_1 als auch in U_2 liegt, und damit auch in $U_1 \cap U_2$ und $U_1 + U_2$. Wir müssen also nur noch die Abgeschlossenheit zeigen.

(a) Es seien $\lambda \in K$ und $x, y \in U_1 \cap U_2$, also $x, y \in U_1$ und $x, y \in U_2$. Da U_1 und U_2 Unterräume sind, liegen damit sowohl $x + y$ als auch λx in U_1 und U_2, d. h. in $U_1 + U_2$.

(b) Es seien $\lambda \in K$ und $x, y \in U_1 + U_2$, also $x = x_1 + x_2$ und $y = y_1 + y_2$ mit $x_1, y_1 \in U_1$ und $x_2, y_2 \in U_2$. Dann gilt wegen der Abgeschlossenheit von U_1 und U_2

$$x + y = x_1 + x_2 + y_1 + y_2 = \underbrace{(x_1 + y_1)}_{\in U_1} + \underbrace{(x_2 + y_2)}_{\in U_2} \in U_1 + U_2,$$

und analog auch

$$\lambda x = \lambda(x_1 + x_2) = \underbrace{\lambda x_1}_{\in U_1} + \underbrace{\lambda x_2}_{\in U_2} \in U_1 + U_2. \qquad \square$$

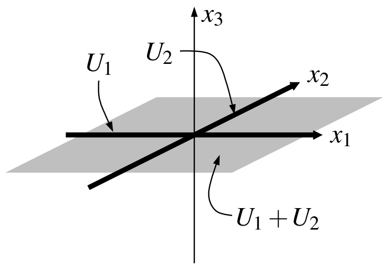
Beispiel 15.12. Für die Unterräume

$$U_1 = \operatorname{Lin}(e_1) \quad \text{und} \quad U_2 = \operatorname{Lin}(e_2)$$

von $\mathbb{R}^3$ ist ihr Durchschnitt natürlich gleich $U_1 \cap U_2 = \{0\}$, und ihre Summe wie im Bild rechts dargestellt die (x_1, x_2)-Ebene

$$U_1 + U_2 = \operatorname{Lin}(e_1, e_2).$$

Aber auch für beliebige Unterräume von K^m können wir ihren Durchschnitt und ihre Summe mit unseren Methoden aus Kapitel 14 konkret berechnen:



Algorithmus 15.13 (Berechnung von Durchschnitten und Summen in K^m)**.** Es seien U_1 und U_2 zwei Unterräume von K^m, die wir nach Satz 15.9 als $U_1 = \operatorname{Lin}(x_1, \dots, x_k)$ und $U_2 = \operatorname{Lin}(y_1, \dots, y_l)$ schreiben können. Die gewählten Erzeuger müssen dabei im Folgenden nicht linear unabhängig sein.

(a) Die Vektoren im Schnitt $U_1 \cap U_2$ erhält man offensichtlich durch Gleichsetzen

$$\lambda_1 x_1 + \cdots + \lambda_k x_k = -\mu_1 y_1 - \cdots - \mu_l y_l \tag{1}$$

der Elemente von U_1 und U_2, und damit durch Auflösen des linearen Gleichungssystems

$$\lambda_1 x_1 + \cdots + \lambda_k x_k + \mu_1 y_1 + \cdots + \mu_l y_l = 0 \tag{2}$$

(die Vorzeichen spielen hierbei keine Rolle, da mit $y \in U_2$ auch $-y \in U_2$ liegt, und sind daher so gewählt, dass sie sich im resultierenden Gleichungssystem wegheben). Die sich als Lösung ergebenden Werte für $\lambda_1, \dots, \lambda_k$ können dann in die linke Seite von (1) eingesetzt werden und liefern die gesuchten Vektoren im Schnitt.

Als konkretes Beispiel betrachten wir die Unterräume

$$U_1 = \operatorname{Lin}\left(\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\1\end{pmatrix}\right) \quad \text{und} \quad U_2 = \operatorname{Lin}\left(\begin{pmatrix}0\\1\\2\end{pmatrix}, \begin{pmatrix}1\\1\\3\end{pmatrix}\right)$$

von $\mathbb{R}^3$. Der Ansatz (2) führt zum folgenden Gleichungssystem in $\lambda_1, \lambda_2, \mu_1, \mu_2$:

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \overset{Z3-Z2\to Z3}{\longrightarrow} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix} \overset{Z2-Z3\to Z2}{\longrightarrow} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{pmatrix}.$$

Der Lösungsraum dieses Gleichungssystems wird also (z. B. nach Bemerkung 14.26 (a)) von $(\lambda_1, \lambda_2, \mu_1, \mu_2) = (1, -1, 2, -1)$ erzeugt. Einsetzen von λ_1 und λ_2 (oder alternativ μ_1 und μ_2) in (1) zeigt also, dass $U_1 \cap U_2$ erzeugt wird von

$$\lambda_1 x_1 + \lambda_2 x_2 = 1 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - 1 \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}.$$

(b) Sind die Unterräume dagegen wie in Bemerkung 15.10 (c) als Kerne von Matrizen

$$U_1 = \operatorname{Ker} A_1 = \{x \in K^m : A_1 x = 0\} \quad \text{und} \quad U_2 = \operatorname{Ker} A_2 = \{x \in K^m : A_2 x = 0\}$$

gegeben, so können wir ihren Durchschnitt offensichtlich sofort hinschreiben als

$$U_1 \cap U_2 = \{x \in K^m : A_1 x = A_2 x = 0\} = \operatorname{Ker} \begin{pmatrix} A_1 \\ \hline A_2 \end{pmatrix}.$$

(c) Die Summe $U_1 + U_2$ ist nach Definition natürlich gegeben durch die Linearkombinationen aller Erzeuger zusammen; es gilt also

$$U_1 + U_2 = \operatorname{Lin}(x_1, \dots, x_k, y_1, \dots, y_l).$$

Beachte dabei, dass die Vektoren $x_1, \dots, x_k, y_1, \dots, y_l$ nicht zusammen linear unabhängig sein müssen – wenn linear unabhängige Erzeuger gesucht sind, müssen wir hierauf noch das Verfahren aus Bemerkung 15.10 (a) anwenden.

Aufgabe 15.14. Es seien U_1 und U_2 Unterräume eines K-Vektorraums V. Zeige, dass $U_1 \cup U_2$ genau dann ein Unterraum von V ist, wenn $U_1 \subset U_2$ oder $U_2 \subset U_1$.

Aufgabe 15.15. Es seien

$$U_1 = \{f : \mathbb{R} \to \mathbb{R} : f(-x) = f(x) \text{ für alle } x \in \mathbb{R}\}$$
$$\text{und} \quad U_2 = \{f : \mathbb{R} \to \mathbb{R} : f(-x) = -f(x) \text{ für alle } x \in \mathbb{R}\}$$

in $\operatorname{Abb}(\mathbb{R}, \mathbb{R})$ die Mengen aller sogenannten geraden bzw. ungeraden Funktionen. Man zeige:

(a) U_1 und U_2 sind Unterräume von $\operatorname{Abb}(\mathbb{R}, \mathbb{R})$.

(b) $U_1 \cap U_2 = \{0\}$ und $U_1 + U_2 = \operatorname{Abb}(\mathbb{R}, \mathbb{R})$.

15.C Lineare Abbildungen

Wie schon am Anfang des letzten Abschnitts angekündigt, wollen wir uns nun mit Abbildungen zwischen Vektorräumen beschäftigen, die mit der Vektorraumstruktur verträglich sind.

Definition 15.16 (Lineare Abbildungen bzw. Morphismen)**.** Es seien V und W zwei Vektorräume über demselben Grundkörper K. Man nennt eine Abbildung $f : V \to W$ eine **lineare Abbildung** (oder **Morphismus** oder **(Vektorraum-)Homomorphismus**), wenn für alle $x, y \in V$ und $\lambda \in K$ gilt, dass

$$f(x+y) = f(x) + f(y) \quad \text{(„}f\text{ ist verträglich mit der Vektoraddition“)}$$
$$\text{und} \quad f(\lambda x) = \lambda f(x) \quad \text{(„}f\text{ ist verträglich mit der Skalarmultiplikation“).}$$

Die Menge aller solchen Morphismen mit Startraum V und Zielraum W wird mit $\operatorname{Hom}_K(V, W)$ bezeichnet (oder auch nur mit $\operatorname{Hom}(V, W)$, wenn der Grundkörper aus dem Zusammenhang klar ist).

Ist $V = K^n$, so schreiben wir statt $f\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right)$ der Einfachheit halber oft nur $f\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

Bemerkung 15.17. Setzt man $x = 0$ und $\lambda = 0$ in Definition 15.16 ein, so erhält man sofort, dass $f(0) = 0$ für jeden Morphismus $f\colon V \to W$ gilt. Für ein festes $a \in V\backslash\{0\}$ ist also z. B. die Verschiebeabbildung $f\colon V \to V,\ x \mapsto x+a$ wegen $f(0) = a \neq 0$ nie ein Morphismus.

Beispiel 15.18.

(a) Für beliebige K-Vektorräume V und W ist die Nullabbildung $f\colon V \to W,\ x \mapsto 0$ natürlich immer ein Morphismus.

(b) Für eine Matrix $A \in K^{m\times n}$ bezeichnen wir mit f_A die Abbildung

$$f_A\colon K^n \times K^m,\ x \mapsto Ax.$$

Sie ist ein Morphismus, denn für alle $x, y \in K^n$ und $\lambda \in K$ gilt

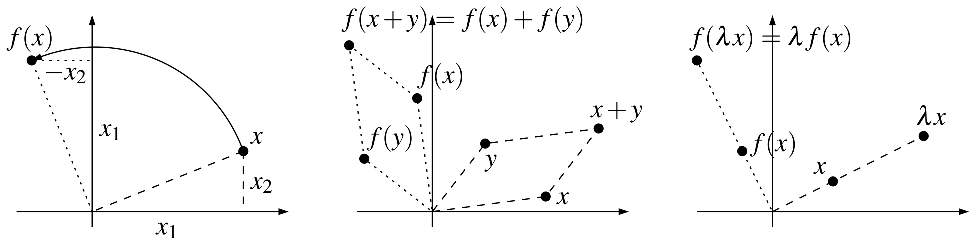
$$f_A(x+y) = A(x+y) = Ax+Ay = f_A(x)+f_A(y) \quad \text{und} \quad f_A(\lambda x) = A(\lambda x) = \lambda(Ax) = \lambda f_A(x)$$

(wie wir bereits in Bemerkung 13.13 festgestellt hatten).

Konkret erhalten wir z. B. für die Matrix

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2} \quad \text{die lineare Abbildung} \quad f_A\colon \mathbb{R}^2 \to \mathbb{R}^2,\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}.$$

Geometrisch beschreibt sie wie im Bild unten links eine Vierteldrehung um den Ursprung. An den anderen beiden Bildern kann man die Morphismuseigenschaft auch gut anschaulich ablesen: Im mittleren Bild ist z. B. das gepunktete Parallelogramm aus der Drehung des gestrichelten entstanden, und der äußerste Punkt ergibt sich damit sowohl durch Addition der Punkte $f(x)$ und $f(y)$ als auch durch Drehung von $x+y$, d. h. es ist $f(x)+f(y) = f(x+y)$. Entsprechendes gilt für die Skalarmultiplikation im rechten Bild.



Anschaulich ist damit auch schon erkennbar, dass Drehungen um andere Winkel (um den Ursprung) ebenfalls Morphismen sein sollten. Wir werden solche allgemeinen Drehungen später in Beispiel **??** und Abschnitt **??** untersuchen.

(c) Die Abbildung

$$f\colon \mathbb{R}^2 \to \mathbb{R},\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto x_1^2 + x_2$$

ist nicht linear, denn es ist z. B.

$$f\begin{pmatrix} -1 \\ 0 \end{pmatrix} + f\begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1+1 = 2 \neq 0 = f\begin{pmatrix} 0 \\ 0 \end{pmatrix} = f\left(\begin{pmatrix} -1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right).$$

(d) Lemma 13.7 besagt genau, dass für alle $m, n \in \mathbb{N}$ die Transposition von Matrizen

$$f\colon K^{m\times n} \to K^{n\times m},\ A \mapsto A^{\mathsf{T}}$$

eine lineare Abbildung ist.

(e) Es sei $V \subset \mathrm{Abb}(\mathbb{R},\mathbb{R})$ der Vektorraum aller reellen Polynomfunktionen aus Beispiel 15.8 (e). Wir betrachten die Abbildung $f\colon V \to V,\ \varphi \mapsto \varphi'$, die jedem Polynom $\varphi\colon x \mapsto \sum_{k=0}^n a_k x^k$ seine *Ableitung*

$$\varphi'\colon x \mapsto \sum_{k=1}^n k a_k x^{k-1} \qquad (*)$$

zuordnet.

Wenn ihr die Ableitung bereits aus der Analysis kennt, wisst ihr aus den Regeln in Beispiel **?? ??** schon, dass die Ableitung eines Polynoms durch $(*)$ gegeben ist, und dass f eine lineare Abbildung ist, da nach Satz **?? ??** und Beispiel **?? ??** für alle differenzierbaren Funktionen φ und ψ sowie $\lambda \in \mathbb{R}$ sowohl $(\varphi+\psi)' = \varphi' + \psi'$ als auch $(\lambda\varphi)' = \lambda\varphi'$ gilt.

Wenn ihr die Ableitung aus der Analysis noch nicht kennt, könnt ihr $(*)$ für die Zwecke der linearen Algebra einfach als *Definition* der Ableitung eines Polynoms ansehen. Man rechnet dann schnell nach, dass f wirklich eine lineare Abbildung ist: Für zwei Polynome $\varphi(x) = \sum_{k=0}^n a_k x^k$ und $\psi(x) = \sum_{k=0}^n b_k x^k$ (wobei n der größere der beiden Grade ist, so dass sowohl φ als auch ψ so geschrieben werden können) sowie $\lambda \in \mathbb{R}$ ist

$$(\varphi+\psi)'(x) = \sum_{k=1}^n k\,(a_k+b_k)x^{k-1} = \sum_{k=1}^n k\,a_k x^{k-1} + \sum_{k=1}^n k\,b_k x^{k-1} = \varphi'(x) + \psi'(x)$$

$$\text{und}\quad (\lambda\varphi)'(x) = \sum_{k=1}^n k\,(\lambda a_k)x^{k-1} = \lambda \sum_{k=1}^n k\,a_k x^{k-1} = \lambda\varphi'(x).$$

Sind der Startraum K^n und der Zielraum K^m für gewisse $m,n \in \mathbb{N}$, so ist die Situation ähnlich wie bei Unterräumen in Satz 15.9 wieder deutlich schöner: In diesem Fall wollen wir jetzt zeigen, dass *jeder* Morphismus $f\colon K^m \to K^n$ von der Form f_A wie in Beispiel 15.18 (b) ist – und zwar sogar mit einer eindeutig bestimmten Matrix $A \in K^{m\times n}$.

Satz und Definition 15.19 (Lineare Abbildungen $K^n \to K^m$ und Abbildungsmatrizen)**.** *Zu jeder linearen Abbildung $f\colon K^n \to K^m$ gibt es genau eine Matrix $A \in K^{m\times n}$ mit*

$$f(x) = Ax \quad \textit{für alle } x \in K^n,$$

*nämlich $A = (f(e_1) \mid \cdots \mid f(e_n))$. Wir nennen sie die **Abbildungsmatrix** von f und bezeichnen sie mit A_f.*

Beweis. Die geforderte Bedingung legt nach Beispiel 13.10 (d) für alle $j = 1,\dots,n$ die j-te Spalte von A fest zu $Ae_j = f(e_j)$. Damit ist A eindeutig bestimmt als $(f(e_1) \mid \cdots \mid f(e_n))$. Da f linear ist, folgt aus dieser Beziehung für die Einheitsvektoren aber auch für alle $x \in K^n$ mit Koordinaten $x_1,\dots,x_n$

$$f(x) = f(x_1e_1+\cdots+x_ne_n) = x_1f(e_1)+\cdots+x_nf(e_n) = x_1Ae_1+\cdots+x_nAe_n = Ax. \qquad \square$$

Wir wollen nun einige elementare Eigenschaften von Morphismen zeigen und beginnen damit, dass Bilder und Urbilder (im Sinne von Definition 2.11) von Unterräumen unter Morphismen immer wieder Unterräume sind.

Lemma 15.20 (Bilder und Urbilder von Unterräumen)**.** *Es sei $f\colon V \to W$ eine lineare Abbildung. Dann gilt:*

(a) *Ist U ein Unterraum von V, so ist $f(U)$ ein Unterraum von W.*

(b) *Ist U ein Unterraum von W, so ist $f^{-1}(U)$ ein Unterraum von V.*

Beweis. Wir müssen das Unterraumkriterium aus Satz 15.6 überprüfen.

(a) Wegen $0 \in U$ ist nach Bemerkung 15.7 zunächst $0 = f(0) \in f(U)$, also ist $f(U) \neq \emptyset$. Wir zeigen nun die Abgeschlossenheit von $f(U)$ bezüglich der Vektoraddition. Es seien dazu $x,y \in f(U)$, d. h. $x = f(u)$ und $y = f(v)$ für gewisse $u,v \in U$. Dann ist auch $u+v \in U$, und damit folgt $x+y = f(u)+f(v) = f(u+v) \in f(U)$.

Genauso zeigt man die Abgeschlossenheit unter der Skalarmultiplikation.

(b) Wegen $f(0) = 0 \in U$ ist zunächst einmal $0 \in f^{-1}(U)$, d. h. es ist $f^{-1}(U) \neq \emptyset$. Wir zeigen jetzt die Abgeschlossenheit von $f^{-1}(U)$ unter der Vektoraddition. Dazu seien $x,y \in f^{-1}(U)$, d. h. $x,y \in V$ mit $f(x), f(y) \in U$. Dann ist auch $f(x+y) = f(x)+f(y) \in U$, also $x+y \in f^{-1}(U)$.

Analog ergibt sich die Abgeschlossenheit unter der Skalarmultiplikation. $\square$

Die wichtigsten Spezialfälle dieses Lemmas sind die folgenden, die wir im Zusammenhang mit Matrizen schon kennengelernt hatten:

Definition 15.21 (Bild und Kern eines Morphismus)**.** Es sei $f \colon V \to W$ ein lineare Abbildung von K-Vektorräumen.

(a) Die Menge $\operatorname{Im} f := f(V) = \{f(x) : x \in V\}$ heißt das **Bild** von f.

(b) Die Menge $\operatorname{Ker} f := f^{-1}(\{0\}) = \{x \in V : f(x) = 0\}$ heißt der **Kern** von f.

Nach Lemma 15.20 gilt offensichtlich $\operatorname{Im} f \leq W$ und $\operatorname{Ker} f \leq V$.

Beispiel 15.22. Im Fall $V = K^n$ und $W = K^m$ sind Bild und Kern eines Morphismus $f \colon K^n \to K^m$ dasselbe wie Bild und Kern der zugehörigen Abbildungsmatrix $A := A_f$ aus Definition 15.19: Da f von der Form $f(x) = Ax$ ist, ist nach Definition 14.19

$$\operatorname{Im} A = \{Ax : x \in K^n\} = \{f(x) : x \in K^n\} = \operatorname{Im} f$$

$$\text{und} \quad \operatorname{Ker} A = \{x \in K^n : Ax = 0\} = \{x \in K^n : f(x) = 0\} = \operatorname{Ker} f.$$

Offensichtlich ist eine lineare Abbildung $f \colon V \to W$ nach Definition genau dann surjektiv, wenn $\operatorname{Im} f = W$. Wir wollen jetzt ein analoges Kriterium auch für die Injektivität zeigen.

Lemma 15.23. *Eine lineare Abbildung $f \colon V \to W$ ist genau dann injektiv, wenn* $\operatorname{Ker} f = \{0\}$.

Beweis.

„$\Rightarrow$“ Ist f injektiv, so hat der Nullvektor höchstens ein Urbild unter f. Wegen $f(0) = 0$ ist das Urbild des Nullvektors also genau der Nullvektor, d. h. es ist $\operatorname{Ker} f = \{0\}$.

„$\Leftarrow$“ Es sei $\operatorname{Ker} f = \{0\}$. Weiterhin seien $x, y \in V$ mit $f(x) = f(y)$. Wegen der Linearität von f gilt dann $f(x-y) = f(x) - f(y) = 0$, mit $\operatorname{Ker} f = \{0\}$ also $x - y = 0$. Damit folgt $x = y$, d. h. f ist injektiv. □

Lemma 15.24 (Umkehrabbildungen und Verkettungen)**.** *Es sei $f \colon V \to W$ ein Morphismus von K-Vektorräumen. Dann gilt:*

(a) *Ist f bijektiv, so ist auch die Umkehrabbildung f^{-1} ein Morphismus.*

(b) *Ist $g \colon W \to Z$ ein weiterer Morphismus von K-Vektorräumen, so ist auch $g \circ f \colon V \to Z$ ein Morphismus.*

Beweis.

(a) Es seien $x, y \in W$; wir setzen $u = f^{-1}(x)$ und $v = f^{-1}(y)$, also $x = f(u)$ und $y = f(v)$. Dann gilt

$$\begin{aligned} f^{-1}(x+y) &= f^{-1}(f(u) + f(v)) \\ &= f^{-1}(f(u+v)) && (f \text{ ist ein Morphismus}) \\ &= u + v && (f^{-1} \text{ ist Umkehrabbildung von } f) \\ &= f^{-1}(x) + f^{-1}(y). \end{aligned}$$

Analog zeigt man die Verträglichkeit mit der Skalarmultiplikation.

(b) Für $x, y \in V$ gilt

$$\begin{aligned} g(f(x+y)) &= g(f(x) + f(y)) && (f \text{ ist ein Morphismus}) \\ &= g(f(x)) + g(f(y)) && (g \text{ ist ein Morphismus}). \end{aligned}$$

Genauso ergibt sich die Verträglichkeit mit der Skalarmultiplikation. □

Im Fall von Vektorräumen der Form K^n lassen sich die Ergebnisse aus Lemma 15.23 und 15.24 wie erwartet auch wieder in die Sprache der Matrizen übersetzen:

Folgerung 15.25. *Es seien $f \colon K^n \to K^m$ ein Morphismus und $A := A_f$ seine Abbildungsmatrix. Dann gilt:*

(a) *Die Abbildung f ist genau dann injektiv, wenn* $\operatorname{rk} A = n$*.*

Die Abbildung f ist genau dann surjektiv, wenn $\operatorname{rk} A = m$*.*

Insbesondere ist f also genau dann bijektiv, wenn A quadratisch und invertierbar ist. In diesem Fall ist die Umkehrabbildung f^{-1} der Morphismus mit Abbildungsmatrix A^{-1}.

(b) *Ist $g\colon K^m \to K^p$ ein weiterer Morphismus, so ist*

$$A_{g\circ f} = A_g \cdot A_f,$$

d. h. die Verkettung von Morphismen entspricht der Multiplikation der zugehörigen Abbildungsmatrizen.

Beweis.

(a) Da f gegeben ist durch $f(x) = Ax$ für alle $x \in K^n$, ist dies genau die Aussage über die universelle Lösbarkeit des Gleichungssystems $Ax = b$ aus Folgerung 14.28 bzw. Bemerkung 14.29.

(b) Mit $B = A_g$ gilt für alle $x \in K^n$

$$(g \circ f)(x) = g(f(x)) = B(Ax) = (BA)x,$$

d. h. $BA = A_g \cdot A_f$ ist die Abbildungsmatrix von $g \circ f$. □

Aufgabe 15.26. Man zeige: Sind $f, g\colon V \to W$ zwei lineare Abbildungen, so gilt

(a) $\operatorname{Ker} f \cap \operatorname{Ker} g \subset \operatorname{Ker}(f+g)$;

(b) $\operatorname{Im}(f+g) \subset \operatorname{Im} f + \operatorname{Im} g$.

Weiterhin gebe man in beiden Fällen ein Beispiel an, das zeigt, dass man im Allgemeinen nicht „$\subset$“ durch „$=$“ ersetzen kann.

Aufgabe 15.27. Es sei $f\colon V \to W$ eine lineare Abbildung zwischen K-Vektorräumen. Ferner sei U ein Unterraum von V mit $U \cap \operatorname{Ker} f = \{0\}$ und $U + \operatorname{Ker} f = V$.

Zeige, dass die Abbildung $f|_U\colon U \to \operatorname{Im} f$ bijektiv ist.

Aufgabe 15.28. Für ein gegebenes $n \in \mathbb{N}_{>0}$ sei $V \le \operatorname{Abb}(\mathbb{R}, \mathbb{R})$ der Unterraum aller Polynomfunktionen vom Grad kleiner als n. Zeige, dass die Abbildung

$$f\colon V \to \mathbb{R}^n,\ \varphi \mapsto (\varphi(1), \dots, \varphi(n)).$$

linear ist, und bestimme Kern und Bild von f.

Aufgabe 15.29. Es sei V ein $\mathbb{R}$-Vektorraum. Eine lineare Abbildung $f\colon V \to V$ heißt *Projektion*, wenn $f \circ f = f$.

(a) Gib für den Fall $V = \mathbb{R}^2$ ein Beispiel für eine Projektion an, bei der sowohl $\operatorname{Ker} f$ als auch $\operatorname{Im} f$ nicht-triviale Unterräume von V sind.

(b) Man zeige: Ist $f\colon V \to V$ eine Projektion, so gilt $\operatorname{Ker} f \cap \operatorname{Im} f = \{0\}$ und $\operatorname{Ker} f + \operatorname{Im} f = V$.

(c) Es seien $f, g\colon V \to V$ zwei Projektionen. Beweise, dass $f + g$ genau dann ebenfalls eine Projektion ist, wenn $\operatorname{Im} f \subset \operatorname{Ker} g$ und $\operatorname{Im} g \subset \operatorname{Ker} f$.

Gilt diese Aussage auch für Vektorräume über einem beliebigen Körper?

Wie wir nun zum Abschluss dieses Kapitels noch sehen wollen, haben bijektive Morphismen wie in Lemma 15.24 (a) in der Praxis eine besondere Bedeutung. Sie haben daher auch einen besonderen Namen:

Definition 15.30 (Isomorphismen)**.** Es seien V und W zwei K-Vektorräume.

(a) Einen bijektiven Morphismus $f\colon V \to W$ (der nach Lemma 15.24 (a) also einen Umkehrmorphismus $f^{-1}\colon W \to V$ besitzt) bezeichnet man als **(Vektorraum-)Isomorphismus**.

(b) V und W heißen **isomorph** (in Zeichen: $V \cong W$), wenn es einen Isomorphismus $f\colon V \to W$ zwischen ihnen gibt.

Beispiel 15.31. Anschaulich bedeutet ein Isomorphismus f zwischen zwei Vektorräumen V und W, dass diese beiden Räume „als Vektorräume ununterscheidbar" sind: Die Objekte in V und W sind zwar unterschiedlich benannt, aber in allen Rechnungen können wir jederzeit mit der bijektiven Abbildung f bzw. der inversen Abbildung f^{-1} zwischen den beiden Darstellungen in V und W hin- und herwechseln, ohne das Endergebnis zu ändern. Die folgenden Beispiele verdeutlichen dies.

(a) Der Unterraum

$$V = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} \in \mathbb{R}^3 : x_1, x_2 \in \mathbb{R} \right\} \leq \mathbb{R}^3 \quad \text{ist mit} \quad f \colon V \to \mathbb{R}^2,\ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

isomorph zu $\mathbb{R}^2$. In der Tat ist in diesem Beispiel offensichtlich, dass f linear und bijektiv ist. Auch anschaulich ist in diesem Fall klar, dass V und $\mathbb{R}^2$ „im Prinzip ununterscheidbar" sind, denn beide Räume sind einfach die reelle Ebene – die im Fall von V lediglich als Koordinatenebene in den $\mathbb{R}^3$ eingebettet ist.

(b) Für alle $m,n \in \mathbb{N}$ ist der Raum $K^{m\times n}$ aller $m \times n$-Matrizen über K isomorph zu K^{mn}: Ein Isomorphismus $f \colon K^{m\times n} \to K^{mn}$ ist einfach dadurch gegeben, dass man die Einträge einer Matrix von links oben nach rechts unten nun untereinander in einen Vektor in K^{mn} schreibt. Auch hier ist klar, dass f linear und bijektiv ist, also dass die Anordnung der Zahlen – einmal als rechteckiges Schema und einmal untereinander geschrieben – nichts an der Vektorraumstruktur ändert, da Addition und Skalarmultiplikation in beiden Fällen einfach komponentenweise ausgeführt werden. (Dass es in $K^{m\times n}$ eine Matrixmultiplikation gibt, in K^{mn} jedoch nicht, spielt hierbei keine Rolle, da dies nicht Teil der Vektorraumaxiome ist.)

(c) Für den Vektorraum V aller reellen Polynome vom Grad höchstens 2 ist

$$f \colon V \to \mathbb{R}^3,\ a_0 + a_1 x + a_2 x^2 \mapsto \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}$$

ein Isomorphismus, d. h. es gilt $V \cong \mathbb{R}^3$. Auch hier ist wieder klar, dass f linear ist, und dass der Vektor der Koeffizienten a_0, a_1, a_2 dieselben Informationen enthält wie das Polynom $a_0 + a_1 x + a_2 x^2$ (siehe Bemerkung 3.22).

(d) Für $m,n \in \mathbb{N}$ mit $m \neq n$ sind K^m und K^n nicht isomorph, da es dann nach Folgerung 15.25 (a) keinen Isomorphismus von K^m nach K^n gibt. So ist anschaulich gesprochen z. B. eine Gerade K^1 „etwas anderes" als eine Ebene K^2. Diese Idee wird in **??** zum zentralen Konzept der Dimension eines Vektorraums führen.

In der Tat hatten wir in Satz 15.19 noch einen weiteren Fall gesehen, in dem zwei Objekte „im Prinzip dasselbe" waren, nämlich lineare Abbildungen von K^n nach K^m und Matrizen in $K^{m\times n}$. Auch diese Aussage können wir jetzt wie folgt mit Isomorphismen exakt formulieren.

Lemma 15.32 (Hom(V,W) als Vektorraum)**.** *Es seien V und W zwei K-Vektorräume.*

(a) Hom(V,W) *ist ein Unterraum von* Abb(V,W) *(und damit also selbst wieder ein K-Vektorraum).*

(b) *Im Fall $V = K^n$ und $W = K^m$ mit $m,n \in \mathbb{N}$ ist*

$$K^{m\times n} \to \operatorname{Hom}(K^n, K^m),\ A \mapsto f_A$$

ein Isomorphismus mit Umkehrabbildung

$$\operatorname{Hom}(K^n, K^m) \to K^{m\times m},\ f \mapsto A_f.$$

Beweis.

(a) Natürlich liegt die Nullabbildung in Hom(V,W). Zur Überprüfung der Abgeschlossenheit von Hom(V,W) bezüglich der Addition seien $f,g \in \operatorname{Hom}(V,W)$, d. h. f und g seien lineare

Abbildungen. Dann gilt für ihre Summe $f+g$ für alle $x \in V$ und $\lambda \in K$

$$\begin{aligned}(f+g)(x+y) &= f(x+y)+g(x+y) && \text{(Definition von } f+g\text{)}\\ &= f(x)+f(y)+g(x)+g(y) && (f \text{ und } g \text{ sind linear})\\ &= (f+g)(x)+(f+g)(y) && \text{(Definition von } f+g\text{)}\end{aligned}$$

sowie analog

$$(f+g)(\lambda x) = f(\lambda x)+g(\lambda x) = \lambda f(x)+\lambda g(x) = \lambda(f+g)(x).$$

Also ist dann auch $f+g$ linear, d. h. es ist $f+g \in \operatorname{Hom}(V,W)$. Analog überprüft man die Abgeschlossenheit von $\operatorname{Hom}(V,W)$ bezüglich der Skalarmultiplikation.

(b) Die Abbildung $K^{m\times n} \to \operatorname{Hom}(K^n,K^m)$, $A \mapsto f_A$ ist linear, denn für alle $A,B \in K^{m\times n}$ sowie $x \in K^n$ und $\lambda \in K$ gilt

$$f_{A+B}(x) = (A+B)x = Ax+Bx = f_A(x)+f_B(x) \quad \text{und} \quad f_{\lambda A}(x) = (\lambda A)x = \lambda(Ax) = \lambda f_A(x),$$

und damit $f_{A+B} = f_A + f_B$ und $f_{\lambda A} = \lambda f_A$. Die Bijektivität dieser Abbildung haben wir bereits in Satz 15.19 gesehen. □

16. Basen und Dimension

Im letzten Kapitel haben wir viele Konzepte der allgemeinen Vektorraumtheorie eingeführt: Neben dem Vektorraumbegriff selbst haben wir Unterräume sowie deren Durchschnitte und Summen betrachtet, Morphismen mit ihrem Bild und Kern definiert, und dazu Umkehrmorphismen sowie Verkettungen untersucht. Dabei haben wir immer wieder gesehen, dass die betrachteten Konstruktionen im Fall von Vektorräumen der Form K^n für ein $n \in \mathbb{N}$ auch ganz explizit beschrieben und berechnet werden können.

Wir wollen nun sehen, wie wir diese Rechenmethoden auf einen beliebigen Vektorraum V übertragen können. Die Idee dafür ist, eine sogenannte *Basis* von V zu finden – dies sind (im Moment der Einfachheit halber endlich viele) Vektoren $x_1, \dots, x_n \in V$, so dass sich jedes $x \in V$ auf eindeutige Art als Linearkombination

$$x = \lambda_1 x_1 + \cdots + \lambda_n x_n$$

dieser ausgewählten Vektoren mit geeigneten Skalaren $\lambda_1, \dots, \lambda_n \in K$ schreiben lässt. Die $\lambda_1, \dots, \lambda_n$ lassen sich dann als „Koordinaten“ von x bezüglich der Vektoren $x_1, \dots, x_n$ auffassen. Da sie das Element $x \in V$ eindeutig beschreiben und einen Vektor in K^n bilden, ist die Abbildung $V \to K^n$, die jedem Vektor seine Koordinaten bezüglich $x_1, \dots, x_n$ zuordnet, ein Isomorphismus. Gemäß Beispiel 15.31 bedeutet dies dann, dass wir mit diesem Isomorphismus alle Rechenmethoden für K^n auf V übertragen können.

16.A Linearkombinationen und Basen

Um die oben beschriebene Idee umzusetzen, müssen wir als Erstes den Begriff der Linearkombination, den wir in Definition 14.21 ja schon für Vektoren in K^n eingeführt hatten, auf beliebige Vektorräume übertragen. Dies funktioniert prinzipiell genau wie erwartet und hat auch die gleiche anschauliche Bedeutung – allerdings müssen wir für den allgemeinen Fall auch unendliche Familien von Vektoren zulassen, aus denen wir die Linearkombinationen bilden.

Definition 16.1 (Familien und Linearkombinationen)**.** Es sei V ein K-Vektorraum.

(a) Eine **Familie** von Vektoren in V ist eine Indexmenge I zusammen mit einem Vektor $x_i \in V$ für alle $i \in I$. Wir schreiben eine solche Familie als $B = (x_i)_{i \in I}$ und nennen die x_i ihre **Elemente**. Ist die Indexmenge I endlich, so sprechen wir auch von einer **endlichen Familie** B, wählen als Indexmenge in der Regel $I = \{1, \dots, n\}$, und schreiben die Familie dann einfach als $B = (x_1, \dots, x_n)$.

(b) Es sei $B = (x_i)_{i \in I}$ eine Familie von Vektoren in V. Eine **Linearkombination** der Vektoren aus B ist ein Vektor der Form

$$\sum_{i \in I} \lambda_i x_i \quad \in V, \tag{$*$}$$

wobei $\lambda_i \in K$ für alle $i \in I$ gilt *und nur endlich viele dieser* λ_i *ungleich* 0 *sind* (so dass der Ausdruck $(*)$ nach Weglassen aller Summanden, die 0 sind, also in jedem Fall nur eine endliche Summe ist). Im Fall der endlichen Indexmenge $I = \{1, \dots, n\}$ schreibt man die Linearkombination in der Regel als $\sum_{i=1}^n \lambda_i x_i$ oder $\lambda_1 x_1 + \cdots + \lambda_n x_n$.

Analog zur bisherigen Notation bezeichnen wir die Menge aller Linearkombinationen von Vektoren aus B mit $\operatorname{Lin} B$.

Bemerkung 16.2.

(a) Beachte, dass eine *Familie* $B = (x_i)_{i \in I}$ nicht das gleiche ist wie eine *Menge* $\{x_i : i \in I\}$ von Vektoren: Im Gegensatz zu einer Menge sind die gegebenen Vektoren in einer Familie den

Elementen einer Indexmenge zugeordnet, haben also z. B. im Fall einer endlichen Familie mit Indexmenge $\{1,\dots,n\}$ eine festgelegte Reihenfolge $x_1,\dots,x_n$. Für die obige Definition 16.1 (b) ist dies noch nicht wichtig (wir hätten $\operatorname{Lin} B$ genauso gut auf die gleiche Art für eine *Menge* B definieren können, und in der Tat wird das in der Literatur auch oft getan), aber später werden wir die gegebene Reihenfolge der Vektoren zwingend benötigen, um die Koordinaten eines Vektors den Elementen von B zuordnen zu können.

(b) Im Folgenden werden wir uns in vielen Fällen nur mit endlichen Familien beschäftigen. In diesem Fall ist die Bedingung in Definition 16.1 (b), dass nur endlich viele der Koeffizienten λ_i ungleich 0 sein dürfen, natürlich automatisch erfüllt und kann daher auch weggelassen werden. Im Fall unendlicher Familien ist sie jedoch sehr wichtig, da der Ausdruck $\sum_{i\in I}\lambda_i x_i$ sonst eine unendliche Summe wäre und damit ohne Konvergenzbetrachtungen wie in Kapitel **??** (die nicht Gegenstand der linearen Algebra sind und in einem allgemeinen Körper auch gar nicht definiert werden könnten) keinen Sinn ergeben würde.

(c) Genau wie in Beispiel 15.8 (b) zeigt man wieder, dass $\operatorname{Lin} B$ für jede Familie B in einem Vektorraum V ein Unterraum ist. Wir nennen $\operatorname{Lin} B$ daher auch hier den von B **erzeugten** oder **aufgespannten Unterraum** von V.

Beispiel 16.3.

(a) Für Vektoren $x_1,\dots,x_n$ in K^m mit $m,n\in\mathbb{N}$ stimmt unsere neue Definition 16.1 (b) von $\operatorname{Lin}(x_1,\dots,x_n)$ offensichtlich mit der alten aus Definition 14.21 (a) überein.

(b) Im Vektorraum $V = \operatorname{Abb}(\mathbb{N},\mathbb{R})$ aller reellen Zahlenfolgen aus Beispiel 15.3 (d) betrachten wir die Familie $B = (e_i)_{i\in\mathbb{N}}$ aller „Einheitsfolgen“ $e_i = (0,\dots,0,1,0,0,\dots)$, wobei die 1 jeweils an der Stelle i steht. Wegen der Endlichkeitsbedingung in Definition 16.1 (b) ist $\operatorname{Lin} B$ dann *nicht* der gesamte Raum aller Folgen, sondern nur die Menge aller Folgen, bei denen nur endlich viele Folgenglieder ungleich 0 sind.

Mit unseren Vorarbeiten können wir nun den Begriff der Basis eines Vektorraums definieren. Dazu müssen wir auch den Begriff der linearen Unabhängigkeit aus Definition 14.21 (b) auf unsere neue Situation erweitern, was wieder genau wie erwartet funktioniert.

Definition 16.4 (Basen von Vektorräumen)**.** Es sei $B = (x_i)_{i\in I}$ eine Familie von Vektoren in einem K-Vektorraum V.

(a) Die Familie B heißt ein **Erzeugendensystem** von V, wenn $\operatorname{Lin} B = V$ gilt, d. h. wenn sich jeder Vektor $x\in V$ als Linearkombination der Vektoren in B schreiben lässt.

(b) Die Familie B heißt **linear abhängig**, wenn es eine Linearkombination $\sum_{i\in I}\lambda_i x_i = 0$ des Nullvektors gibt, in der mindestens ein λ_i ungleich 0 ist (man nennt dies auch eine *nicht-triviale Linearkombination* des Nullvektors).

Ist das Gegenteil der Fall, folgt aus der Linearkombination $\sum_{i\in I}\lambda_i x_i = 0$ des Nullvektors mit zunächst beliebigen $\lambda_i\in K$ also bereits, dass alle λ_i gleich 0 sein müssen, so heißt B **linear unabhängig**.

(c) Die Familie B heißt eine **Basis** von V, wenn B ein Erzeugendensystem von V und linear unabhängig ist.

Beispiel 16.5.

(a) Sind $x_1,\dots,x_n$ in K^m linear unabhängig, so ist $(x_1,\dots,x_n)$ nach Definition eine Basis von $\operatorname{Lin}(x_1,\dots,x_n)$. In der Tat können wir unsere Konvention in Bemerkung 15.10, dass wir unter der „Berechnung“ eines Unterraums $U\le K^m$ die Angabe linear unabhängiger $x_1,\dots,x_n$ mit $U=\operatorname{Lin}(x_1,\dots,x_n)$ verstehen wollen, jetzt so umformulieren, dass wir eine Basis von U angeben wollen.

(b) Nach (a) ist insbesondere $(e_1,\dots,e_n)$ eine Basis von $K^n = \operatorname{Lin}(e_1,\dots,e_n)$, da die Einheitsvektoren nach Beispiel 14.22 (a) linear unabhängig sind. Sie wird die **Standardbasis** von

K^n genannt und ist sicher die wichtigste und einfachste Basis von K^n – aber bei weitem nicht die einzige, wie wir gleich in Folgerung 16.6 sehen werden.

Als Spezialfall davon für $n = 0$ ist die leere Familie eine Basis des Nullvektorraums.

(c) Im Vektorraum $V \leq \mathrm{Abb}(\mathbb{R},\mathbb{R})$ aller Polynomfunktionen aus Beispiel 15.8 (d) ist die Familie $B = (x^i)_{i\in\mathbb{N}}$ aller Potenzfunktionen mit natürlichen Exponenten eine Basis von V:

- B ist ein Erzeugendensystem, da jede Polynomfunktion nach Definition eine (endliche!) Linearkombination der Potenzfunktionen x^i ist;
- B ist linear unabhängig, da eine nicht-triviale Linearkombination der x^i nach dem Koeffizientenvergleich aus Lemma 3.21 nie die Nullfunktion sein kann.

Genauso ist natürlich $(x^0, x^1, \dots, x^n)$ eine Basis des Vektorraums aller Polynomfunktionen vom Grad höchstens n.

(d) Umformuliert bedeutet Beispiel 16.3 (b), dass die Einheitsfolgen $e_i = (0,\dots,0,1,0,0,\dots)$, wobei die 1 an Position i steht, kein Erzeugendensystem des Folgenraums $\mathrm{Abb}(\mathbb{N},\mathbb{R})$ bilden. Sie sind jedoch linear unabhängig, denn ist $\sum_{i\in\mathbb{N}} \lambda_i e_i = (0,0,0,\dots)$ (für eine gewisse Summe mit nur endlich vielen $\lambda_i \neq 0$), so folgt durch Vergleich des i-ten Folgengliedes natürlich sofort $\lambda_i = 0$ für alle i.

Wie ihr sicher schon erwartet, können wir auch die in Definition 16.4 eingeführten Eigenschaften im Fall von Familien im Vektorraum K^n wieder ganz einfach explizit überprüfen – für die lineare Unabhängigkeit hatten wir dies ja auch schon in Folgerung 14.30 (a) gesehen.

Folgerung 16.6 (**Rangkriterium** für Basen in K^m). *Es seien $m,n \in \mathbb{N}$ und $x_1,\dots,x_n \in K^m$; wir setzen $A := (x_1 \,|\, \cdots \,|\, x_n) \in K^{m\times n}$.*

(a) *Die Familie $(x_1,\dots,x_n)$ ist genau dann linear unabhängig, wenn $\mathrm{rk}\,A = n$ gilt.*

(b) *Die Familie $(x_1,\dots,x_n)$ erzeugt genau dann K^m, wenn $\mathrm{rk}\,A = m$ gilt.*

Insbesondere ist $(x_1,\dots,x_n)$ also genau dann eine Basis von K^m, wenn $\mathrm{rk}\,A = m = n$ ist.

Beweis.

(a) ist genau unser altes Rangkriterium aus Folgerung 14.30.

(b) Die Familie $(x_1,\dots,x_n)$ erzeugt genau dann K^m, wenn $\mathrm{Im}\,A \overset{14.21\,(a)}{=} \mathrm{Lin}(x_1,\dots,x_n) = K^m$ ist, also wenn die Abbildung f_A surjektiv ist. Dies ist nach Folgerung 15.25 (a) genau für $\mathrm{rk}\,A = m$ der Fall. □

Bemerkung 16.7.

(a) Nach Folgerung 16.6 hat jede Basis von K^m genau m Elemente.

(b) Haben wir umgekehrt genau m Vektoren in K^m, so besagt Folgerung 16.6 ebenfalls, dass diese Vektoren schon eine Basis von K^m bilden, wenn wir sie nur als linear unabhängig *oder* Erzeugendensystem voraussetzen. Dies halbiert also den Arbeitsaufwand, wenn wir von gegebenen Vektoren in K^m überprüfen wollen, ob sie eine Basis bilden.

In der Tat können wir im Fall des Vektorraums K^m nicht nur zeigen, dass jede Basis genau m Elemente hat, sondern auch Erzeugendensysteme mit mehr als m bzw. linear unabhängige Familien mit weniger als m Elementen konstruktiv zu einer Basis machen:

Folgerung 16.8 (**Basisauswahl** und **Basisergänzung** in K^m). *Es sei $(x_1,\dots,x_n)$ eine Familie von n Vektoren in K^m.*

(a) *Ist $(x_1,\dots,x_n)$ ein Erzeugendensystem von K^m, so ist $n \geq m$, und man kann aus diesen Vektoren m Vektoren auswählen, die eine Basis von K^m bilden.*

Mit anderen Worten ist eine Basis von K^m dasselbe wie ein minimales *Erzeugendensystem von K^m – also ein Erzeugendensystem, das sich nicht weiter verkleinern lässt.*

(b) *Ist $(x_1,\dots,x_n)$ linear unabhängig, so ist $n \le m$, und man kann diese Vektoren mit geeigneten Einheitsvektoren zu m Vektoren ergänzen, die eine Basis von K^m bilden.*

Mit anderen Worten ist eine Basis von K^m dasselbe wie eine maximale linear unabhängige Familie in K^m – also eine linear unabhängige Familie, die sich nicht weiter vergrößern lässt.

Beweis.

(a) Mit dem Verfahren aus Folgerung 14.31 **??** können wir aus den gegebenen Vektoren linear unabhängige auswählen, die ebenfalls K^m erzeugen, also eine Basis von K^m sind. Nach Bemerkung 16.7 (a) müssen dies dann genau m Vektoren sein.

(b) Mit dem Algorithmus aus Folgerung **??** können wir die gegebenen Vektoren zu m linear unabhängigen Vektoren ergänzen, die nach Bemerkung 16.7 (b) dann eine Basis von K^m sein müssen. □

16.B Endlich erzeugte Vektorräume

Um Vektorräume (wie in der Einleitung zu diesem Kapitel beschrieben) mit Hilfe von Basen untersuchen zu können, müssen wir uns jetzt natürlich noch fragen, ob denn überhaupt jeder Vektorraum eine Basis besitzt – aus den Vektorraumaxiomen ist das ja nicht offensichtlich. In der Tat ist das aber so: Man kann beweisen, dass jeder beliebige Vektorraum eine Basis hat.

Wir wollen uns beim Beweis dieser Tatsache allerdings auf Vektorräume beschränken, die von *endlich vielen* Elementen erzeugt werden können und dann auch eine *endliche* Basis besitzen (den allgemeinen Beweis könnt ihr z. B. in [GK] Proposition II.2.22 finden). Dies liegt zum einen daran, dass der Beweis für Vektorräume mit unendlichen Basen deutlich komplizierter und abstrakter ist. Zum anderen – und das ist fast der wichtigere Grund – ist der Beweis im allgemeinen Fall *nicht konstruktiv* und daher eigentlich nur von theoretischem Interesse. Betrachten wir z. B. noch einmal den $\mathbb{R}$-Vektorraum $\mathrm{Abb}(\mathbb{N},\mathbb{R})$ aller reellen Zahlenfolgen aus Beispiel 15.3 (d) und Beispiel 16.5 (d): Wir werden in Beispiel 16.13 **??** sehen, dass dieser Vektorraum keine *endliche* Basis besitzt. Man weiß nun zwar aufgrund des oben angegebenen Satzes, dass dieser Folgenraum dennoch eine (unendliche) Basis hat, aber niemand kann eine solche Basis konkret angeben! Versucht doch einmal, eine Basis zu finden – ihr werdet sehr schnell merken, dass das aussichtslos ist. Zur Erinnerung: Ihr müsstet dazu eine (unendliche) Familie von Folgen hinschreiben, so dass *jede beliebige* Folge auf *eindeutige* Art eine *endliche* Linearkombination der Folgen ist, die ihr ausgewählt habt.

Formal bedeutet dies, dass wir uns in Zukunft in der Regel auf Vektorräume mit der folgenden Zusatzbedingung beschränken wollen.

Definition 16.9 (Endlich erzeugte Vektorräume)**.** Ein K-Vektorraum V heißt **endlich erzeugt**, wenn er ein Erzeugendensystem aus endlich vielen Vektoren besitzt.

Analog zur Basisauswahl in Folgerung 16.8 (a) würde man nun sicher erwarten, dass man aus einem solchen endlichen Erzeugendensystem eine Basis auswählen und so insbesondere die Existenz einer Basis zeigen kann. In der Tat werden wir dies nun beweisen – allerdings können wir dazu nicht unseren alten Beweis verwenden, da er auf Rechnungen mit Matrizen beruht und damit nicht auf abstrakte Vektorräume übertragbar ist. Stattdessen benutzen wir hier ein anderes (und ebenfalls konstruktives) Verfahren zur Basisauswahl, das iterativ so lange Vektoren aus dem gegebenen Erzeugendensystem streicht, bis eine Basis vorliegt.

Satz 16.10 (Existenz von Basen)**.** *Jeder endlich erzeugte Vektorraum hat eine endliche Basis.*

Beweis. Es sei V ein endlich erzeugter Vektorraum, d. h. es existiert ein endliches Erzeugendensystem $B = (x_1,\dots,x_n)$ von V.

Ist B nun bereits linear unabhängig, so sind wir natürlich fertig. Andernfalls gibt es eine nichttriviale Linearkombination $\lambda_1 x_1 + \cdots + \lambda_n x_n = 0$ des Nullvektors, d. h. es gibt ein $i \in \{1,\dots,n\}$ mit $\lambda_i \neq 0$. Wir setzen nun $B' = (x_1,\dots,x_{i-1},x_{i+1},\dots,x_n)$ und behaupten, dass B' immer noch ein Erzeugendensystem von V ist.

In der Tat ist dies leicht einzusehen: Aus der Linearkombination des Nullvektors folgt wegen $\lambda_i \neq 0$

$$x_i = \frac{1}{\lambda_i} \cdot \left(-\lambda_1 x_1 - \cdots - \lambda_{i-1} x_{i-1} - \lambda_{i+1} x_{i+1} - \cdots - \lambda_n x_n \right) \quad \in \operatorname{Lin} B'.$$

Damit enthält der Unterraum $\operatorname{Lin} B'$ alle Vektoren $x_1, \dots, x_n$, und somit auch alle Linearkombinationen davon, d. h. es ist bereits $\operatorname{Lin} B' = \operatorname{Lin} B = V$.

Ist B' nun linear unabhängig, so sind wir wieder fertig. Andernfalls wiederholen wir das obige Verfahren so lange, bis die resultierende Familie linear unabhängig und damit eine Basis von V ist (dies muss spätestens nach n Schritten passieren, da dann keine Vektoren mehr übrig sind und die leere Familie natürlich linear unabhängig ist). □

Mit einer solchen Basis $(x_1, \dots, x_n)$ eines Vektorraums V erhalten wir nun zu jedem $x \in V$ zugehörige Koordinaten, die einen Isomorphismus von V nach K^n liefern:

Konstruktion 16.11 (Koordinatenabbildungen)**.** Es sei V ein Vektorraum mit einer gegebenen Basis $B = (x_1, \dots, x_n)$. Dann ist die Abbildung

$$f\colon K^n \to V, \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \mapsto \lambda_1 x_1 + \cdots + \lambda_n x_n$$

offensichtlich linear, und außerdem bijektiv:

- Sie ist surjektiv, weil B ein Erzeugendensystem von V ist; und
- sie ist nach Lemma 15.23 injektiv, da aus der linearen Unabhängigkeit von B folgt, dass $\operatorname{Ker} f = \{0\}$ ist.

Jeder Vektor $x \in V$ lässt sich also auf eindeutige Art als Linearkombination $x = \lambda_1 x_1 + \cdots + \lambda_n x_n$ schreiben. Wir nennen $\lambda_1, \dots, \lambda_n$ die **Koordinaten** von x bezüglich B, und die Umkehrabbildung von f

$$\Phi_B\colon V \to K^n, \ \lambda_1 x_1 + \cdots + \lambda_n x_n \mapsto \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

die **Koordinatenabbildung** zu B. Mit diesem Isomorphismus Φ_B ist also $V \cong K^n$.

Entscheidend ist dabei die Beobachtung, dass die Vektorräume K^n für verschiedene n nach Beispiel 15.31 (d) ja nicht isomorph zueinander sind und die Zahl n in Konstruktion 16.11 damit durch V schon eindeutig bestimmt ist. Dies führt zum folgenden wichtigen Begriff der Dimension eines Vektorraums.

Satz und Definition 16.12 (Dimension von Vektorräumen)**.** *Alle Basen eines endlich erzeugten Vektorraums V (sind endlich und) haben gleich viele Elemente. Man nennt diese Anzahl Elemente die **Dimension** von V, und schreibt sie als $\dim_K V$ oder einfach $\dim V$.*

*Ist V nicht endlich erzeugt (und hat damit natürlich auch keine endliche Basis), so schreiben wir formal $\dim V = \infty$. Ein endlich erzeugter Vektorraum wird daher oft auch als **endlich-dimensionaler Vektorraum** bezeichnet.*

Beweis. Nach Satz 16.10 hat V eine endliche Basis $B = (x_1, \dots, x_n)$; dies liefert nach Konstruktion 16.11 einen Isomorphismus $\Phi_B\colon V \to K^n$.

Es sei nun C eine weitere Basis von V. Wäre C unendlich, so gäbe es insbesondere $n+1$ linear unabhängige Vektoren $y_1, \dots, y_{n+1} \in V$, und damit eine Koordinatenabbildung $\Phi_C\colon U \to K^{n+1}$ mit $U := \operatorname{Lin}(y_1, \dots, y_{n+1}) \leq V$. Dann wäre aber die Abbildung $f\colon K^{n+1} \to K^n$, $x \mapsto \Phi_B(\Phi_C^{-1}(x))$ linear und injektiv; ihre Abbildungsmatrix $A_f \in K^{n \times (n+1)}$ hätte also nach Folgerung 15.25 (a) den Rang $n+1$ – was offensichtlich ein Widerspruch zu Bemerkung 14.9 (a) ist.

Also ist $C = (y_1, \dots, y_m)$ endlich, und liefert eine Koordinatenabbildung $\Phi_C \colon V \to K^m$. Damit ist aber $\Phi_B \circ \Phi_C^{-1} \colon K^m \to K^n$ ein Isomorphismus, was nach Beispiel 15.31 (d) nur für $m = n$ möglich ist. □

Beispiel 16.13.

(a) Da die Standardbasis von K^n für ein $n \in \mathbb{N}$ genau n Elemente hat, ist natürlich $\dim K^n = n$.

(b) Nach Beispiel 16.5 (c) hat der Vektorraum V aller reellen Polynomfunktionen vom Grad höchstens n eine Basis $B = (x^0, x^1, \dots, x^n)$. Es ist also $\dim V = n+1$, und die zu B gehörige Koordinatenabbildung ist

$$\Phi_B \colon V \to \mathbb{R}^{n+1},\ a_0 + a_1 x + \cdots + a_n x^n \mapsto \begin{pmatrix} a_0 \\ \vdots \\ a_n \end{pmatrix}$$

wie in Beispiel 15.31 (c).

(c) Es sei $V = \mathrm{Abb}(\mathbb{N}, \mathbb{R})$ der Raum aller reellen Zahlenfolgen aus Beispiel 15.3 (d). Nach Beispiel 16.5 (d) besitzt V eine unendliche linear unabhängige Familie von Vektoren und kann somit wie im Beweis von Satz 16.12 nicht endlich erzeugt sein. Es ist also $\dim V = \infty$.

Bemerkung 16.14 (Klassifikation endlich-dimensionaler Vektorräume)**.** Fassen wir die Ergebnisse aus Satz 16.10, Konstruktion 16.11 und Satz 16.12 zusammen, so haben wir insgesamt gezeigt:

Jeder endlich erzeugte Vektorraum V ist (über die Koordinatenabbildung zu einer beliebigen Basis) isomorph zu K^n für genau ein $n \in \mathbb{N}$, nämlich für $n = \dim V$.

Dies ist ein sehr wichtiges Resultat: Immer wenn man eine neue mathematische Struktur eingeführt und ein paar Beispiele untersucht hat, wäre es natürlich wünschenswert, wenn man vielleicht sogar eine *vollständige* Liste aller Beispiele angeben könnte – in unserem Fall also eine vollständige Liste aller (endlich erzeugten) Vektorräume. Dabei soll „vollständig“ immer „vollständig bis auf Isomorphie“ bedeuten, da wir ja in Beispiel 15.31 schon gesehen haben, dass isomorphe Vektorräume von ihrer Struktur her ohnehin ununterscheidbar sind, so dass es uns bei isomorphen Vektorräumen natürlich reichen sollte, wenn einer von ihnen in unserer Liste steht.

Bei vielen mathematischen Strukturen ist eine derartige Klassifikation schlichtweg aussichtslos, weil es viel zu viele Beispiele gibt, die auch keinem ersichtlichen Schema folgen. Dies ist z. B. bei Gruppen (siehe Definition 3.1) der Fall – niemand kann eine vollständige Liste aller Gruppen (bis auf Isomorphie) angeben. Es ist eine Besonderheit der linearen Algebra, dass dies bei endlich erzeugten Vektorräumen anders ist: Wie wir jetzt gesehen haben, sind diese Vektorräume genau durch ihre Dimension klassifiziert, d. h. zu jedem Körper K und jeder natürlichen Zahl $n \in \mathbb{N}$ gibt es bis auf Isomorphie *genau einen* K-Vektorraum dieser Dimension, nämlich K^n.

Auf diese Art sollten wir nun also alle unsere Ergebnisse und Rechenverfahren zum Vektorraum K^n auf einen beliebigen endlich erzeugten Vektorraum V übertragen können, indem wir statt mit den Elementen $x \in V$ mit den Koordinatenvektoren $\Phi_B(x) \in K^n$ zu einer beliebigen Basis B rechnen. Dies werden wir im Rest dieses Kapitels ausführlich tun.

Wir beginnen damit, dass Isomorphismen (wie z. B. Koordinatenabbildungen) wie erwartet alle in diesem Kapitel eingeführten Begriffe erhalten.

Lemma 16.15. *Es seien $f \colon V \to W$ ein Morphismus zwischen K-Vektorräumen, $B = (x_1, \dots, x_n)$ eine endliche Familie in V und $f(B) := (f(x_1), \dots, f(x_n))$ die zugehörige Familie der Bildvektoren in W. Dann gilt:*

(a) *Ist f surjektiv und B ein Erzeugendensystem von V, so ist $f(B)$ ein Erzeugendensystem von W.*

(b) *Ist f injektiv und B linear unabhängig, so ist auch $f(B)$ linear unabhängig.*

Insbesondere erhalten Isomorphismen also Erzeugendensysteme, die lineare Unabhängigkeit, und damit auch Basen und die Dimension.

Beweis.

(a) Es sei $y \in W$ beliebig. Da f surjektiv ist, gibt es ein $x \in V$ mit $f(x) = y$, und weil B ein Erzeugendensystem von V ist, können wir $x = \lambda_1 x_1 + \cdots + \lambda_n x_n$ für gewisse $\lambda_1, \dots, \lambda_n \in K$ schreiben. Damit ist aber auch

$$y = f(x) = f(\lambda_1 x_1 + \cdots + \lambda_n x_n) = \lambda_1 f(x_1) + \cdots + \lambda_n f(x_n) \quad \in \operatorname{Lin} f(B).$$

Also ist $\operatorname{Lin} f(B) = W$, d. h. $f(B)$ ist ein Erzeugendensystem von W.

(b) Es seien $\lambda_1, \dots, \lambda_n \in K$ mit $\lambda_1 f(x_1) + \cdots + \lambda_n f(x_n) = 0$, also (weil f ein Morphismus ist) mit

$$f(\lambda_1 x_1 + \cdots + \lambda_n x_n) = 0.$$

Nach Voraussetzung ist f injektiv, d. h. es ist $\operatorname{Ker} f = \{0\}$ aufgrund von Lemma 15.23. Also folgt bereits $\lambda_1 x_1 + \cdots + \lambda_n x_n = 0$ und damit auch $\lambda_1 = \cdots = \lambda_n = 0$, da B linear unabhängig ist. Die Familie $f(B)$ ist somit linear unabhängig. □

Beispiel 16.16 (Basen und Dimension von Matrixräumen)**.** Wir hatten in Beispiel 15.31 (b) schon gesehen, dass der Matrixraum $K^{m \times n}$ isomorph zu K^{mn} ist, indem man alle Einträge der Matrix untereinander schreibt. Insbesondere ist also $\dim K^{m \times n} = mn$. Außerdem bilden die Matrizen, die bezüglich dieses Isomorphismus den Einheitsvektoren in K^{mn} entsprechen, nach Lemma 16.15 eine Basis von $K^{m \times n}$ – nämlich die Matrizen, in denen jeweils ein Eintrag 1 ist und alle anderen 0. So ist z. B.

$$\left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

eine Basis von $K^{2 \times 2}$.

Folgerung 16.17 (Dimension von Produkten)**.** *Sind V und W zwei endlich-dimensionale Vektorräume, so gilt*

$$\dim(V \times W) = \dim V + \dim W.$$

Eine analoge Aussage gilt natürlich auch für Produkte von mehr als zwei Vektorräumen.

Beweis. Nach Bemerkung 16.14 gibt es Isomorphismen $f\colon V \to K^n$ und $g\colon W \to K^m$ mit $n = \dim V$ und $m = \dim W$ (nämlich Koordinatenabbildungen zu beliebigen Basen). Dann ist aber auch

$$V \times W \to K^n \times K^m \cong K^{n+m},\ (x, x') \mapsto (f(x), g(x'))$$

ein Isomorphismus (mit Umkehrabbildung $(y, y') \mapsto (f^{-1}(y), g^{-1}(y'))$), und damit folgt aus Lemma 16.15

$$\dim(V \times W) = n + m = \dim V + \dim W.$$ □

Folgerung 16.18 (**Basisauswahl** und **Basisergänzung**)**.** *Es sei B eine Familie von n Vektoren in einem endlich-dimensionalen Vektorraum V.*

(a) *Ist $n = \dim V$ und B ein Erzeugendensystem von V oder linear unabhängig, so ist B bereits eine Basis von V.*

(b) *Ist B ein Erzeugendensystem von V, so ist $n \geq \dim V$, und man kann aus B eine Basis von V auswählen.*

Mit anderen Worten ist eine Basis von V dasselbe wie ein minimales Erzeugendensystem *von V.*

(c) *Ist B linear unabhängig, so ist $n \leq \dim V$, und man kann B mit geeigneten Vektoren einer fest gewählten Basis C zu einer Basis von V ergänzen.*

Mit anderen Worten ist eine Basis von V dasselbe wie eine maximale linear unabhängige *Familie in V.*

Beweis. Es seien $m = \dim V$ und $C = (y_1, \dots, y_m)$ eine Basis von V. Alle Aussagen der Folgerung ergeben sich direkt mit dem Isomorphismus $\Phi_C \colon V \to K^m$ aus den entsprechenden Aussagen für K^m in Bemerkung 16.7 (b) bzw. Folgerung 16.8.

Wir zeigen exemplarisch Teil (c): Es sei $B = (x_1, \dots, x_n)$ linear unabhängig in V. Nach Lemma 16.15 (b) ist dann $(\Phi_C(x_1), \dots, \Phi_C(x_n))$ linear unabhängig in K^m. Nach Folgerung 16.8 (b) können wir diese Vektoren nun mit geeigneten Einheitsvektoren $e_{k_1}, \dots, e_{k_{m-n}}$ zu einer Basis von K^m ergänzen. Anwenden von Φ_C^{-1} auf diese Basis liefert dann wiederum nach Lemma 16.15 die Basis $\left(x_1, \dots, x_n, \Phi_C^{-1}(e_{k_1}), \dots, \Phi_C^{-1}(e_{k_{m-n}})\right) = (x_1, \dots, x_n, y_{k_1}, \dots, y_{k_{m-n}})$ von V. □

Folgerung 16.19. *Es sei U ein Unterraum eines endlich-dimensionalen K-Vektorraums V. Dann ist auch U endlich-dimensional, und es ist $\dim U \le \dim V$.*

Gilt sogar $\dim U = \dim V$, so ist $U = V$.

Beweis. Wir wählen wieder einen beliebigen Isomorphismus $f \colon V \to K^n$ mit $n = \dim V$. Dann ist $f(U)$ nach Lemma 15.20 (a) ein Unterraum von K^n, also nach Satz 15.9 endlich erzeugt. Nach Lemma 16.15 ist damit dann auch $f^{-1}(f(U)) = U$ endlich erzeugt.

Nach Folgerung 16.18 können wir damit eine Basis von U zu einer von V ergänzen. Damit folgt sofort $\dim U \le \dim V$, und im Fall der Gleichheit wird bei der Ergänzung kein zusätzlicher Vektor benötigt, so dass dann schon $U = V$ ist. □

Mit diesen Ergebnissen können wir nun schon viele Rechenmethoden von K^n auf beliebige endlich-dimensionale Vektorräume übertragen: z. B. die Bestimmung von Erzeugendensystemen und linearer Unabhängigkeit, die Basisauswahl und -ergänzung, oder die Berechnung von Durchschnitten und Summen von Unterräumen. Hier ist ein konkretes Beispiel dafür:

Beispiel 16.20. Im Vektorraum V aller reellen Polynome vom Grad höchstens 2 seien

$$U_1 = \operatorname{Lin}(1, x+x^2) \quad \text{und} \quad U_2 = \operatorname{Lin}(x+2x^2, 1+x+3x^2).$$

Um mit diesen Unterräumen konkrete Berechnungen durchzuführen, können wir gemäß Beispiel 16.13 (c) die Basis $B = (1, x, x^2)$ von V wählen und die Koordinatenvektoren bezüglich B benutzen. Konkret sind diese Koordinatenvektoren der oben gegebenen Polynome die Vektoren

$$\Phi_B(1) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \Phi_B(x+x^2) = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \Phi_B(x+2x^2) = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \Phi_B(1+x+3x^2) = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} \qquad (*)$$

in $\mathbb{R}^3$. Damit können wir z. B. berechnen:

(a) Da die ersten beiden Vektoren in $(*)$ linear unabhängig in $\mathbb{R}^3$ sind, sind 1 und $x+x^2$ linear unabhängig in U_1, und damit eine Basis von U_1. Es ist also $\dim U_1 = 2$. Analog sieht man auch $\dim U_2 = 2$.

(b) Da die ersten beiden Vektoren in $(*)$ z. B. mit dem zweiten Einheitsvektor e_2 zu einer Basis von $\mathbb{R}^3$ ergänzt werden, ist $\left(1, x+x^2, \Phi_B^{-1}(e_2)\right) = (1, x+x^2, x)$ eine Basis von V.

(c) Um eine Basis des Durchschnitts $U_1 \cap U_2$ zu bestimmen, berechnen wir zunächst das Bild dieses Durchschnitts unter Φ_B. Diese Rechnung haben wir schon in Algorithmus 15.13 (a) durchgeführt: Es ist

$$\Phi_B(U_1) \cap \Phi_B(U_2) = \operatorname{Lin}\left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\right) \cap \operatorname{Lin}\left(\begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}\right) = \operatorname{Lin}\begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} \quad \subset \mathbb{R}^3.$$

Anwenden von Φ_B^{-1} liefert also $U_1 \cap U_2 = \operatorname{Lin}(1-x-x^2)$; insbesondere hat $U_1 \cap U_2$ damit die Dimension 1.

16.C Lineare Abbildungen zwischen endlich erzeugten Vektorräumen

Wir wollen nun lineare Abbildungen zwischen endlich erzeugten Vektorräumen genauer untersuchen. Als Erstes übertragen wir dazu die Definition des Rangs mit Hilfe von Bemerkung 16.13 (b) von Matrizen auf Morphismen:

Definition 16.21 (Rang eines Morphismus)**.** Für eine lineare Abbildung $f\colon V \to W$ zwischen zwei Vektorräumen V und W heißt

$$\operatorname{rk} f := \dim \operatorname{Im} f \quad \in \mathbb{N} \cup \{\infty\}$$

der **Rang** von f.

Analog zum vorherigen Abschnitt können wir solche linearen Abbildungen für endlich erzeugte Vektorräume V und W nun mit Hilfe von Koordinatenabbildungen auf lineare Abbildungen zwischen K^n und K^m mit $n = \dim V$ und $m = \dim W$ zurückführen: Wollen wir eine lineare Abbildung $f\colon V \to W,\ x \mapsto f(x)$ beschreiben, so können wir stattdessen genauso gut Basen B und C von V bzw. W wählen und die zugehörige lineare Abbildung $K^n \to K^m$ betrachten, die den Koordinatenvektor $\Phi_B(x)$ auf $\Phi_C(f(x))$ abbildet. Dieser Morphismus lässt sich dann genau wie bisher durch Multiplikation mit einer Abbildungsmatrix erhalten.

Formal bedeutet dies einfach, dass wir Beispiel 15.18 (b) und den Satz 15.19 über Abbildungsmatrizen mehr oder weniger abschreiben können, wenn wir in den Matrixprodukten alle Vektoren (sowohl im Start- als auch im Zielraum) durch ihre Koordinatenvektoren ersetzen. Dies wollen wir nun machen und den so erhaltenen Zusammenhang zwischen linearen Abbildungen $V \to W$ und Matrizen in $K^{m\times n}$ genauer untersuchen.

Konstruktion 16.22. Es seien V und W zwei endlich-dimensionale Vektorräume mit $n := \dim V$ und $m := \dim W$. Ferner seien B und C fest gewählte Basen von V bzw. W.

Zu jeder Matrix $A \in K^{m\times n}$ ist dann die Abbildung

$$f_A^{B,C}\colon V \to W,\ x \mapsto \Phi_C^{-1}(A \cdot \Phi_B(x))$$

(also mit $f(x) = \Phi_C^{-1}(A \cdot \Phi_B(x))$, d. h. $\Phi_C(f(x)) = A \cdot \Phi_B(x)$ für alle $x \in V$) als Verkettung linearer Abbildungen offensichtlich linear.

Satz und Definition 16.23 (Lineare Abbildungen $V \to W$ und Abbildungsmatrizen)**.** *Es seien V und W zwei endlich-dimensionale Vektorräume mit $n := \dim V$ und $m := \dim W$. Ferner wählen wir Basen $B = (x_1,\dots,x_n)$ und $C = (y_1,\dots,y_m)$ von V bzw. W.*

Dann gibt es zu jeder linearen Abbildung $f\colon V \to W$ genau eine Matrix $A \in K^{m\times n}$ mit $f = f_A^{B,C}$ wie in Konstruktion 16.22, also mit

$$\Phi_C(f(x)) = A \cdot \Phi_B(x) \quad \text{für alle } x \in V,$$

*nämlich $A = (\,\Phi_C(f(x_1)) \mid \cdots \mid \Phi_C(f(x_n))\,)$. Wir nennen sie die **Abbildungsmatrix** von f bezüglich B und C und bezeichnen sie mit $A_f^{B,C}$.*

Beweis. Die geforderte Bedingung legt nach Beispiel 13.10 (d) für alle $j = 1,\dots,n$ die j-te Spalte von A fest zu

$$Ae_j = A \cdot \Phi_B(x_j) = \Phi_C(f(x_j)).$$

Damit ist A eindeutig bestimmt als $(\,\Phi_C(f(x_1)) \mid \cdots \mid \Phi_C(f(x_n))\,)$. Da f linear ist, folgt aus dieser Beziehung für die Einheitsvektoren aber auch für alle $x = \lambda_1 x_1 + \cdots + \lambda_n x_n \in V$ (mit $\lambda_1,\dots,\lambda_n \in K$) mit Koordinatenvektor $\Phi_B(x) = \lambda_1 e_1 + \cdots + \lambda_n e_n$

$$\begin{aligned}\Phi_C(f(x)) &= \Phi_C(f(\lambda_1 x_1 + \cdots + \lambda_n x_n)) = \lambda_1 \Phi_C(f(x_1)) + \cdots + \lambda_n \Phi_C(f(x_n)) \\ &= \lambda_1 A e_1 + \cdots + \lambda_n A e_n = A \cdot \Phi_B(x).\end{aligned}$$

□

Bemerkung 16.24. Es seien wieder V und W zwei endlich-dimensionale Vektorräume der Dimensionen $n := \dim V$ und $m := \dim W$, sowie B und C Basen von V bzw. W.

(a) Ist bereits $V = K^n$ und $W = K^m$, und sind B und C die Standardbasen dieser Vektorräume, so sind die Koordinatenabbildungen Φ_B und Φ_C die Identität auf V bzw. W. Satz 16.23 stimmt dann also genau mit Satz 15.19 überein.

(b) Wie in Lemma 15.32 (b) zeigt man, dass die Abbildung $\operatorname{Hom}(V,W) \to K^{m\times n}$, $f \mapsto A_f^{B,C}$ ein Isomorphismus (mit Umkehrabbildung $K^{m\times n} \to \operatorname{Hom}(V,W)$, $A \mapsto f_A^{B,C}$) ist. Es ist also $\dim \operatorname{Hom}(V,W) = \dim K^{m\times n} = mn = \dim V \cdot \dim W$.

(c) Man kann sich die Satz 16.23 zugrunde liegende Idee auch an dem rechts dargestellten Diagramm verdeutlichen: In der oberen Zeile ist eine lineare Abbildung $f \colon V \to W$ dargestellt. Mit Hilfe der vertikal dargestellten Koordinatenabbildungen (die ja Isomorphismen sind, was durch die Schlangen am Beginn der Pfeile angedeutet werden soll) können wir daraus eine Abbildung $g = \Phi_C \circ f \circ \Phi_B^{-1}$ von K^n nach K^m konstruieren, indem wir im Diagramm „den Umweg über f nehmen“.

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\wr"', "\Phi_B"] & W \arrow[d, "\wr", "\Phi_C"'] \\ K^n \arrow[r, "g"'] & K^m \end{tikzcd}$$

Dieser neue Morphismus $g \colon K^n \to K^m$ ist nun genau die am Anfang dieses Abschnitts beschriebene Abbildung, die für alle $x \in V$ den Koordinatenvektor von x auf den von $f(x)$ abbildet. Wir können sie mit einer gewöhnlichen Abbildungsmatrix A im Sinne von Satz 15.19 beschreiben; es ist also

$$g(v) = Av, \quad \text{d. h.} \quad \Phi_C(f(\Phi_B^{-1}(v))) = Av \quad \text{für alle } v \in K^n.$$

Mit der Substitution $v = \Phi_B(x)$ ist dies nun aber äquivalent zu

$$\Phi_C(f(x)) = A \cdot \Phi_B(x) \quad \text{für alle } x \in V,$$

also genau zu unserer Bedingung aus Satz 16.23.

(d) In der Situation von (c) lassen sich das Bild und der Kern von f leicht aus $\operatorname{Im} g = \operatorname{Im} A$ bzw. $\operatorname{Ker} g = \operatorname{Ker} A$ berechnen und sind damit auch rechnerisch schnell aus der Abbildungsmatrix $A = A_f^{B,C}$ zu bestimmen: Es ist

$$\begin{aligned} & \operatorname{Im} f = f(V) = \Phi_C^{-1}(g(\Phi_B(V))) = \Phi_C^{-1}(g(K^n)) = \Phi_C^{-1}(\operatorname{Im} g) \\ \text{und} \quad & \operatorname{Ker} f = f^{-1}(\{0\}) = \Phi_B^{-1}(g^{-1}(\Phi_C(\{0\}))) = \Phi_B^{-1}(g^{-1}(\{0\})) = \Phi_B^{-1}(\operatorname{Ker} g). \end{aligned}$$

Insbesondere ist damit $\operatorname{rk} f = \dim \operatorname{Im} f = \dim \operatorname{Im} A \overset{16.13\,(b)}{=} \operatorname{rk} A$: Der Rang einer linearen Abbildung $f \colon V \to W$ ist gleich dem Rang seiner Abbildungsmatrix bezüglich beliebiger Basen von V und W. Außerdem erhalten wir so aus Folgerung 15.25 die analoge Aussage für solche linearen Abbildungen

$$\begin{aligned} & f \text{ ist surjektiv} \;\Leftrightarrow\; g \text{ ist surjektiv} \;\Leftrightarrow\; \operatorname{rk} A = m \;\Leftrightarrow\; \operatorname{rk} f = \dim W \\ \text{und} \quad & f \text{ ist injektiv} \;\Leftrightarrow\; g \text{ ist injektiv} \;\Leftrightarrow\; \operatorname{rk} A = n \;\Leftrightarrow\; \operatorname{rk} f = \dim V. \end{aligned}$$

Im Fall $V = W$ ist f also genau dann surjektiv, wenn f injektiv ist.

Beispiel 16.25. Es sei V der Vektorraum aller reellen Polynome vom Grad höchstens 2 mit der Basis $B = (1,x,x^2)$, und W der Vektorraum aller Polynome vom Grad höchstens 1 mit der Basis $C = (1,x)$ (siehe Beispiel 16.5 (c)). Wir betrachten wie in Beispiel 15.18 (e) die lineare Abbildung $f \colon V \to W$, $\varphi \mapsto \varphi'$, die einem Polynom seine Ableitung zuordnet.

(a) Um die Abbildungsmatrix $A_f^{B,C}$ zu bestimmen, müssen wir nach Definition 16.23 die Basisvektoren von B abbilden und als Linearkombinationen der Basiselemente von C schreiben:

$$\begin{aligned} f(1) &= 1' &&= 0 &&= 0 \cdot 1 + 0 \cdot x, \\ f(x) &= x' &&= 1 &&= 1 \cdot 1 + 0 \cdot x, \\ f(x^2) &= (x^2)' &&= 2x &&= 0 \cdot 1 + 2 \cdot x. \end{aligned}$$

Die Koeffizienten dieser Linearkombinationen bilden also die Koordinatenvektoren von $f(1)$, $f(x)$ und $f(x^2)$; wir schreiben sie als Spalten in eine Matrix und erhalten

$$A_f^{B,C} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

(b) Auch die umgekehrte Richtung, aus der Abbildungsmatrix die Abbildung zu rekonstruieren, ist nicht weiter schwierig, wenn man sich daran erinnert, dass die Matrix immer nur Koordinatenvektoren sieht. Angenommen, wir wollen $f(\varphi)$ für $\varphi = 2x^2 + 3x + 4$, also letztlich die Ableitung φ', nur aus der Kenntnis der Abbildungsmatrix $A_f^{B,C}$ bestimmen. Dann brauchen wir zunächst den Koordinatenvektor $\Phi_B(\varphi)$ und können diesen dann an die Abbildungsmatrix multiplizieren: Wegen $\varphi = 4 \cdot 1 + 3 \cdot x + 2 \cdot x^2$ ist

$$\Phi_B(\varphi) = \begin{pmatrix} 4 \\ 3 \\ 2 \end{pmatrix}, \quad \text{und damit} \quad A_f^{B,C} \cdot \Phi_B(\varphi) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}.$$

Nach Satz 16.23 ist dies nun der Koordinatenvektor $\Phi_C(f(\varphi))$ des Bildes $f(\varphi)$. Damit ist $f(\varphi) = 3 \cdot 1 + 4 \cdot x = 4x + 3$ (was in der Tat die Ableitung von φ ist).

(c) Der Kern der Abbildungsmatrix aus (a) ist offensichtlich $\operatorname{Ker} A_f^{B,C} = \operatorname{Lin}(e_1)$. Nach Bemerkung 16.24 (d) ist also $\operatorname{Ker} f = \Phi_B^{-1}(\operatorname{Lin}(e_1)) = \operatorname{Lin}(1)$ der Unterraum, der vom konstanten Polynom 1 erzeugt wird, also genau der Unterraum aller konstanten Polynome. In der Tat sind dies natürlich auch genau die Polynome, deren Ableitung gleich 0 ist.

(d) Nach Bemerkung 16.24 (d) ist $\operatorname{rk} f = \operatorname{rk} A_f^{B,C} = 2$.

Nach Konstruktion hängen unsere gerade eingeführten Abbildungsmatrizen $A_f^{B,C}$ natürlich von der (letztlich willkürlichen) Wahl der Basen B und C im Start- bzw. Zielraum der Abbildung f ab. Wir wollen nun untersuchen, wie sich diese Abbildungsmatrizen ändern, wenn man zu anderen Basen übergeht. Dazu benötigen wir die sogenannten Basiswechselmatrizen, die letztlich ein Spezialfall von Abbildungsmatrizen sind.

Definition 16.26 (Basiswechselmatrizen)**.** Es seien $B = (x_1, \dots, x_n)$ und $B' = (x'_1, \dots, x'_n)$ zwei Basen eines endlich-dimensionalen K-Vektorraums V. Dann heißt die Abbildungsmatrix der Identität id_V bezüglich der Startbasis B' und Zielbasis B, nach Definition 16.23 also

$$A^{B',B} := A_{\mathrm{id}}^{B',B} = \big(\, \Phi_B(x'_1) \mid \cdots \mid \Phi_B(x'_n) \,\big) \quad \in K^{n \times n},$$

die **Basiswechselmatrix** von B' nach B.

Bemerkung 16.27. Es seien B und B' zwei Basen eines endlich-dimensionalen Vektorraums V.

(a) Offensichtlich ist stets $A^{B,B} = E$.

(b) Nach Satz 16.23 ist $A^{B',B}$ die eindeutig bestimmte Matrix mit $\Phi_B(x) = A^{B',B} \cdot \Phi_{B'}(x)$ für alle $x \in V$. Die Basiswechselmatrix wandelt also einfach nur einen Koordinatenvektor bezüglich B' in einen bezüglich B um – was auch ihren Namen erklärt.

Beispiel 16.28. Wollen wir für die beiden Basen

$$B = \left(\begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) \qquad \text{und} \qquad B' = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right)$$

von $\mathbb{R}^2$ die Basiswechselmatrix $A^{B',B}$ bestimmen, so müssen wir die beiden Basisvektoren von B' nach Definition 16.26 als Linearkombination der Vektoren aus B schreiben: Es ist

$$\begin{aligned} \begin{pmatrix} 1 \\ 0 \end{pmatrix} &= \quad 1 \cdot \begin{pmatrix} 0 \\ -1 \end{pmatrix} + 1 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ \text{und} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} &= (-1) \cdot \begin{pmatrix} 0 \\ -1 \end{pmatrix} + 0 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \end{aligned}$$

Die Koeffizienten dieser Linearkombinationen bilden wieder die Koordinatenvektoren bezüglich B, wir schreiben sie also in die Spalten der gesuchten Matrix

$$A^{B',B} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}.$$

Lemma 16.29. *Es sei $B = (x_1, \dots, x_n)$ eine Basis eines endlich erzeugten Vektorraums V. Dann gilt:*

(a) *Für jede weitere Basis B' von V ist $A^{B',B}$ invertierbar mit $(A^{B',B})^{-1} = A^{B,B'}$.*

(b) *Für jede invertierbare Matrix $T \in \mathrm{GL}(n,K)$ gibt es eine Basis B' von V mit $A^{B',B} = T$.*

(c) *Für jede invertierbare Matrix $T \in \mathrm{GL}(n,K)$ gibt es eine Basis B' von V mit $A^{B,B'} = T$.*

Beweis.

(a) Nach Bemerkung 16.27 (b) ist $\Phi_B(x) = A^{B',B} \cdot \Phi_{B'}(x)$ und $\Phi_{B'}(x) = A^{B,B'} \cdot \Phi_B(x)$ für alle $x \in V$. Setzen wir dies ineinander ein, erhalten wir für alle $x \in V$

$$\Phi_B(x) = A^{B',B} \cdot A^{B,B'} \cdot \Phi_B(x).$$

Wiederum nach Bemerkung 16.27 (b) ist $A^{B',B} \cdot A^{B,B'}$ also die Basiswechselmatrix von B nach B, d. h. nach Bemerkung 16.27 (a) gilt $A^{B',B} \cdot A^{B,B'} = E$ und damit $(A^{B',B})^{-1} = A^{B,B'}$.

(b) Wir setzen $x'_i = \Phi_B^{-1}(Te_i)$ für $i = 1, \dots, n$. Da T invertierbar und Φ_B^{-1} ein Isomorphismus ist, ist auch $K^n \to V$, $x \mapsto \Phi_B^{-1}(Tx)$ ein Isomorphismus, und bildet nach Lemma 16.15 damit die Standardbasis $(e_1, \dots, e_n)$ auf eine Basis $B' := (x'_1, \dots, x'_n)$ von V ab.

Für $i = 1, \dots, n$ ist die i-te Spalte der Basiswechselmatrix $A^{B',B}$ nun nach Definition

$$\Phi_B(x'_i) = \Phi_B(\Phi_B^{-1}(Te_i)) = Te_i,$$

also die i-te Spalte von T. Damit ist wie gewünscht $A^{B',B} = T$.

(c) Nach (b) gibt es eine Basis B' mit $A^{B',B} = T^{-1}$, nach (a) also mit $A^{B,B'} = T$. □

Mit diesen Basiswechselmatrizen können wir nun konkret angeben, wie sich Abbildungsmatrizen bei einem Basiswechsel transformieren.

Satz 16.30 (Verhalten von Abbildungsmatrizen unter Basiswechsel)**.** *Es sei $f \colon V \to W$ ein Morphismus zwischen endlich erzeugten Vektorräumen mit gegebenen Basen B bzw. C.*

(a) *Sind B' und C' zwei weitere Basen von V bzw. W, so gilt*

$$A_f^{B',C'} = A^{C,C'} \cdot A_f^{B,C} \cdot A^{B',B}.$$

(b) *Sind umgekehrt $S \in \mathrm{GL}(m,K)$ und $T \in \mathrm{GL}(n,K)$ zwei invertierbare Matrizen, so gibt es Basen B' und C' von V bzw. W, so dass $A^{C,C'} = S$ und $A^{B',B} = T$, und damit*

$$A_f^{B',C'} = S \cdot A_f^{B,C} \cdot T.$$

Beweis.

(a) Nach Definition 16.23 bzw. Bemerkung 16.27 (b) gilt

$$\begin{aligned} \Phi_{C'}(y) &= A^{C,C'} \cdot \Phi_C(y) && \text{für alle } y \in W, \\ \Phi_C(f(x)) &= A_f^{B,C} \cdot \Phi_B(x) && \text{für alle } x \in V, \\ \text{und} \qquad \Phi_B(x) &= A^{B',B} \cdot \Phi_{B'}(x) && \text{für alle } x \in V. \end{aligned}$$

Setzen wir dies für $y = f(x)$ ineinander ein, so erhalten wir

$$\Phi_{C'}(f(x)) = A^{C,C'} \cdot A_f^{B,C} \cdot A^{B',B} \cdot \Phi_{B'}(x) \quad \text{für alle } x \in V,$$

und damit nach Satz 16.23 wie gewünscht $A_f^{B',C'} = A^{C,C'} \cdot A_f^{B,C} \cdot A^{B',B}$.

(b) Nach Lemma 16.29 (b) und (c) existieren Basen B' und C' mit $A^{C,C'} = S$ und $A^{B',B} = T$; die behauptete Formel für die Abbildungsmatrix ergibt sich dann aus (a). □

Bemerkung 16.31 (Normalform von Abbildungsmatrizen)**.** Es seien V und W zwei endlich-dimensionale Vektorräume mit $n = \dim V$ und $m = \dim W$.

Nach Satz 16.30 (b) sind zwei Matrizen $A, A' \in K^{m\times n}$ genau dann eine Abbildungsmatrix für dieselbe Abbildung $f\colon V \to W$, nur bezüglich verschiedener Basen von V und W, wenn es $S \in \mathrm{GL}(m,K)$ und $T \in \mathrm{GL}(n,K)$ gibt mit $A' = SAT$ – also genau dann, wenn A und A' im Sinne von Definition 14.13 zueinander äquivalent sind.

Damit gibt es zum Normalformensatz 14.15 für Matrixäquivalenz auch eine Version für Morphismen: Zu jeder linearen Abbildung $f\colon V \to W$ gibt es Basen B und C von V bzw. W, so dass

$$A_f^{B,C} = \left(\begin{array}{c|c} E_r & 0 \\ \hline 0 & 0 \end{array}\right) \quad \in K^{m\times n},$$

wobei nach Bemerkung 16.24 (d) dann $r = \operatorname{rk} f$ gilt. Analog zu Satz 14.15 nennt man dies auch die **Normalform** der Abbildungsmatrix von f. Wenn es uns also gerade nicht wichtig ist, welche Basen wir für V und W verwenden, können wir stets annehmen, dass die Abbildungsmatrix von f eine solche Normalform hat.

17. Quotientenräume und Dimensionsformeln

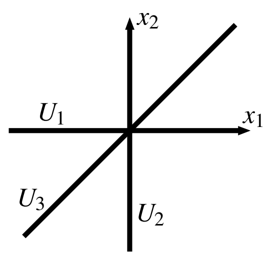
Im letzten Kapitel haben wir den zentralen Begriff der Dimension eines Vektorraums eingeführt. Für alle unsere bisherigen Konstruktionen von Vektorräumen (z. B. als Durchschnitt oder Summe von Unterräumen, als Bild oder Kern einer linearen Abbildung) kennen wir dabei auch explizite Verfahren, wie wir die zugehörigen Dimensionen berechnen können – zunächst mit dem Gauß-Algorithmus für Unterräume von K^n, und dann über Basen bzw. Koordinatenabbildungen auch im allgemeinen Fall.

Es gibt aber auch einige wichtige Beziehungen zwischen diesen Konstruktionen, die es oft ermöglichen, die Dimensionen von bestimmten Vektorräumen auch ohne explizite Rechnung zu bestimmen oder miteinander in Beziehung zu setzen. Wir wollen daher nun in diesem Kapitel noch einige weitere nützliche Konstruktionen und Resultate der Vektorraumtheorie behandeln, mit denen sich unter anderem Dimensionen in vielen Fällen einfach ablesen lassen.

17.A Komplemente und Quotientenräume

Als Erstes wollen wir wie in Lemma 15.11 noch einmal die Konstruktion der Summe $U_1 + \cdots + U_n$ von Unterräumen $U_1, \dots, U_n$ eines Vektorraums V betrachten.

Bemerkung 17.1 (Eindeutigkeit der Summendarstellung)**.** Jeder Vektor in einer Summe $U_1 + \cdots + U_n$ von Unterräumen eines Vektorraums V lässt sich nach Definition als $x_1 + \cdots + x_n$ mit $x_i \in U_i$ für alle $i = 1, \dots, n$ schreiben. Allerdings ist diese Darstellung im Allgemeinen nicht eindeutig: Betrachten wir wie im Bild rechts die drei Ursprungsgeraden



$$U_1 = \text{Lin}\begin{pmatrix}1\\0\end{pmatrix}, \quad U_2 = \text{Lin}\begin{pmatrix}0\\1\end{pmatrix}, \quad U_3 = \text{Lin}\begin{pmatrix}1\\1\end{pmatrix},$$

in $\mathbb{R}^2$, so hat z. B. der Vektor

$$\begin{pmatrix}1\\1\end{pmatrix} = \underbrace{\begin{pmatrix}1\\0\end{pmatrix}}_{\in U_1} + \underbrace{\begin{pmatrix}0\\1\end{pmatrix}}_{\in U_2} + \underbrace{\begin{pmatrix}0\\0\end{pmatrix}}_{\in U_3} = \underbrace{\begin{pmatrix}0\\0\end{pmatrix}}_{\in U_1} + \underbrace{\begin{pmatrix}0\\0\end{pmatrix}}_{\in U_2} + \underbrace{\begin{pmatrix}1\\1\end{pmatrix}}_{\in U_3} \qquad \in U_1 + U_2 + U_3 = \mathbb{R}^2$$

zwei verschiedene Darstellungen dieser Art. Ist die Darstellung jedoch immer eindeutig, so geben wir dieser Situation einen besonderen Namen:

Definition 17.2 (Direkte Summe von Unterräumen)**.** Es seien $U_1, \dots, U_n$ Untervektorräume eines K-Vektorraums V und $U = U_1 + \cdots + U_n$. Hat jedes $x \in U$ eine *eindeutige* Darstellung der Form $x = x_1 + \cdots + x_n$ mit $x_i \in U_i$ für alle $i = 1, \dots, n$, so nennt man die Summe **direkt**. Möchte man dies auch in der Notation andeuten, so schreibt man dafür $U = U_1 \oplus \cdots \oplus U_n$.

Die Summe in Bemerkung 17.1 oben ist also nicht direkt – was natürlich einfach daran liegt, dass die drei aufspannenden Vektoren von U_1, U_2 und U_3 linear abhängig sind. In der Tat kann man sich direkte Summen als eine Verallgemeinerung des Konzepts der linearen Unabhängigkeit auf Unterräume vorstellen.

Lemma 17.3 (Dimension direkter Summen)**.** *Die Summe $U_1 + \cdots + U_n$ von Unterräumen $U_1, \dots, U_n$ eines K-Vektorraums V ist genau dann direkt, wenn die Abbildung*

$$U_1 \times \cdots \times U_n \to U_1 + \cdots + U_n, \ (x_1, \dots, x_n) \mapsto x_1 + \cdots + x_n$$

ein Isomorphismus ist. Ist V endlich-dimensional, so gilt in diesem Fall also insbesondere

$$\dim(U_1 \oplus \cdots \oplus U_n) = \dim U_1 + \cdots + \dim U_n.$$

Beweis. Es ist klar, dass die gegebene Abbildung in jedem Fall linear ist; außerdem ist sie nach Definition der Summe $U_1 + \cdots + U_n$ stets surjektiv. Injektiv ist sie genau dann, wenn für alle $x_i, y_i \in U_i$ aus $x_1 + \cdots + x_n = y_1 + \cdots + y_n$ bereits $(x_1, \dots, x_n) = (y_1, \dots, y_n)$, also $x_i = y_i$ für alle i folgt. Dies bedeutet nach Definition 17.2 aber genau, dass die Summe direkt ist.

Ist V darüber hinaus endlich-dimensional, so gilt dies nach Folgerung 16.19 auch für die Unterräume $U_1, \dots, U_n$. Da endlich erzeugte isomorphe Vektorräume nach Lemma 16.15 die gleiche Dimension haben, ergibt sich aus Folgerung 16.17 also

$$\dim(U_1 \oplus \cdots \oplus U_n) = \dim(U_1 \times \cdots \times U_n) = \dim U_1 + \cdots + \dim U_n. \qquad \square$$

Im Fall von nur zwei Unterräumen kann man besonders einfach feststellen, ob ihre Summe direkt ist:

Lemma 17.4. *Die Summe $U_1 + U_2$ von zwei Unterräumen U_1 und U_2 eines K-Vektorraums V ist genau dann direkt, wenn $U_1 \cap U_2 = \{0\}$.*

Beweis. Nach Lemma 17.3 ist die Summe $U_1 + U_2$ genau dann direkt, wenn die Abbildung

$$f \colon U_1 \times U_2 \to U_1 + U_2, \ (x_1, x_2) \mapsto x_1 + x_2$$

ein Isomorphismus ist. Wir hatten im Beweis dieses Lemmas aber auch schon gesehen, dass f stets linear und surjektiv ist. Also ist die Summe $U_1 + U_2$ genau dann direkt, wenn f injektiv ist, d. h. nach Lemma 15.23 genau dann, wenn $\operatorname{Ker} f = \{(0,0)\}$. Nun ist aber

$$\begin{aligned} \operatorname{Ker} f &= \{(x_1, x_2) : x_1 \in U_1, x_2 \in U_2, x_1 + x_2 = 0\} \\ &= \{(x_1, -x_1) : x_1 \in U_1, -x_1 \in U_2\} \\ &= \{(x_1, -x_1) : x_1 \in U_1 \cap U_2\}, \end{aligned}$$

und damit ist wie behauptet genau dann $\operatorname{Ker} f = \{(0,0)\}$, wenn $U_1 \cap U_2 = \{0\}$. $\square$

Beispiel 17.5.

(a) Die Summe $U_1 + U_2$ der x_1-Achse und der x_2-Achse in $\mathbb{R}^3$ in Beispiel 15.12 ist direkt, denn in diesem Fall ist natürlich $U_1 \cap U_2 = \{0\}$. In der Tat sieht man in diesem Beispiel auch sofort, dass sich jeder Vektor in der x_1-x_2-Ebene $U_1 + U_2$ eindeutig als Summe von einem Vektor in U_1 und einem in U_2 schreiben lässt.

(b) Die Summe $U_1 + U_2 + U_3$ in Bemerkung 17.1 ist hingegen nicht direkt, wie wir dort bereits gesehen hatten. Allerdings ist in diesem Fall trotzdem $U_1 \cap U_2 \cap U_3 = \{0\}$ – was zeigt, dass sich die Aussage von Lemma 17.4 nicht genauso auf mehr als zwei Summanden übertragen lässt. Die zu Lemma 17.4 analoge Aussage für allgemeine Summen ist stattdessen die folgende:

Aufgabe 17.6. Zeige, dass die folgenden Aussagen für Unterräume $U_1, \dots, U_n$ eines K-Vektorraums V äquivalent sind:

(a) $U = U_1 \oplus \cdots \oplus U_n$;

(b) $U = U_1 + \cdots + U_n$ und $U_i \cap (U_1 + \cdots + U_{i-1} + U_{i+1} + \cdots + U_n) = \{0\}$ für alle $i = 1, \dots, n$.

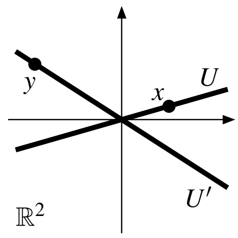
Einen besonderen Namen hat die Situation, wenn die Summe zweier Unterräume direkt und gleich dem gesamten Vektorraum ist.

Definition 17.7 (Komplemente)**.** Es sei U ein Unterraum eines K-Vektorraums V. Ein Unterraum $U' \le V$ heißt **Komplement** von U in V, wenn $U \oplus U' = V$ (nach Lemma 17.4 also $U + U' = V$ und $U \cap U' = \{0\}$) gilt.

Bemerkung 17.8 (Dimension von Komplementen)**.** Nach Lemma 17.3 gilt für jedes Komplement U' eines Untervektorraums U in einem endlich-dimensionalen Vektorraum V die Dimensionsformel $\dim U + \dim U' = \dim V$, also $\dim U' = \dim V - \dim U$.

Beispiel 17.9 (Nichteindeutigkeit von Komplementen)**.** Wie im Bild unten rechts seien $U = \operatorname{Lin}(x)$ und $U' = \operatorname{Lin}(y)$ zwei verschiedene Ursprungsgeraden in $\mathbb{R}^2$. Da x und y dann keine Vielfachen voneinander sind, sind diese beiden Vektoren also linear unabhängig; nach Bemerkung 16.7 (b) bilden sie daher eine Basis des zweidimensionalen Vektorraums $\mathbb{R}^2$.

Nach Algorithmus 15.13 (c) ist damit $U + U' = \operatorname{Lin}(x, y) = \mathbb{R}^2$, außerdem gilt offensichtlich $U \cap U' = \{0\}$. Also ist U' ein Komplement von U.



Da es zu einer gegebenen Ursprungsgeraden U in $\mathbb{R}^2$ aber natürlich unendlich viele Geraden $U' \neq U$ gibt, folgt daraus insbesondere, dass Komplemente von Unterräumen in der Regel nicht eindeutig sind. Wir wollen nun aber sehen, dass Komplemente zumindest im endlich-dimensionalen Fall stets existieren:

Satz 17.10 (Existenz von Komplementen)**.** *Jeder Unterraum U eines endlich-dimensionalen K-Vektorraums V besitzt ein Komplement.*

Beweis. Nach Folgerung 16.19 ist U endlich erzeugt, hat also nach Satz 16.10 eine Basis $(x_1, \dots, x_n)$. Wir ergänzen sie gemäß Folgerung 16.18 (c) zu einer Basis $(x_1, \dots, x_n, y_1, \dots, y_m)$ von V und behaupten, dass $U' := \operatorname{Lin}(y_1, \dots, y_m)$ dann ein Komplement von U ist.

- Es ist offensichtlich $U + U' = V$, denn nach Algorithmus 15.13 (c) ist
$$U + U' = \operatorname{Lin}(x_1, \dots, x_n) + \operatorname{Lin}(y_1, \dots, y_m) = \operatorname{Lin}(x_1, \dots, x_n, y_1, \dots, y_m) = V.$$
- Es ist $U \cap U' = \{0\}$: Für $x \in U \cap U'$, also $x = \lambda_1 x_1 + \dots + \lambda_n x_n = \mu_1 y_1 + \dots + \mu_m y_m$ für gewisse $\lambda_1, \dots, \lambda_n, \mu_1, \dots, \mu_m \in K$, erhalten wir durch Subtraktion
$$0 = x - x = \lambda_1 x_1 + \dots + \lambda_n x_n - \mu_1 y_1 - \dots - \mu_m y_m.$$
Da die Familie $(x_1, \dots, x_n, y_1, \dots, y_m)$ linear unabhängig ist, ist dies aber nur möglich für $\lambda_1 = \dots = \lambda_n = \mu_1 = \dots = \mu_m = 0$, also für $x = 0$.

Also ist U' ein Komplement von U in V. □

Bemerkung 17.11 (Berechnung von Komplementen)**.** Beachte, dass der Beweis von Satz 17.10 konstruktiv ist, d. h. auch die konkrete Berechnung eines Komplements ermöglicht: Möchte man ein Komplement U' zu einem Unterraum U eines endlich-dimensionalen Vektorraums V berechnen, muss man nur eine Basis von U zu einer Basis von V ergänzen; die dafür hinzu genommenen Vektoren bilden dann eine Basis eines Komplements U'.

So haben wir z. B. in Beispiel 16.20 (b) im Raum V aller Polynome vom Grad höchstens 2 die Basis $(1, x+x^2)$ von $U := \operatorname{Lin}(1, x+x^2)$ mit dem Polynom x^2 zu einer Basis $(1, x+x^2, x^2)$ von V ergänzt; dementsprechend ist $U' := \operatorname{Lin}(x^2)$ ein Komplement von U in V.

Komplemente von Unterräumen sind in der Praxis sehr nützlich und haben auch eine einfache anschauliche Deutung: Ist U' ein Komplement eines Unterraums U in einem Vektorraum V, so lässt sich ja jeder Vektor aus V nach Definition 17.7 eindeutig als Summe aus einem Vektor in U und einem „Restvektor“ in U' schreiben. Wir können uns U' also in gewissem Sinne als einen Unterraum vorstellen, der diese „Restteile“ von Vektoren in V misst, wenn man ihren Anteil in U heraus nimmt. Unschön ist an dieser Konstruktion allerdings, dass ein Komplement nach Beispiel 17.9 nicht eindeutig bestimmt und damit ein recht unnatürliches Objekt ist. Was im obigen Sinne der Restteil eines Vektors in V nach Herausnehmen des Anteils in U ist, lässt sich also nicht beantworten, solange man nicht eine (letztlich willkürliche) Wahl eines Komplements von U in V getroffen hat.

Wir wollen nun eine deutlich schönere Konstruktion einführen, die solche Restteile auch ohne willkürliche Wahlen messen kann. Der Preis dafür ist, dass der Vektorraum, der diese Restteile auf ganz natürliche Art beschreibt, kein *Unterraum* von V mehr ist, sondern ein sogenannter *Quotientenraum*: ein Raum von Äquivalenzklassen von Vektoren in V wie in Abschnitt 2.B, wobei wir zwei Vektoren in V miteinander identifizieren wollen, wenn sie sich um ein Element von U voneinander unterscheiden. Diejenigen von euch, die auch die Vorlesung „Algebraische Strukturen“ besuchen, kennen diese Idee vielleicht bereits von den Faktorgruppen [G, Kapitel 6].

Lemma und Definition 17.12. *Es seien V ein K-Vektorraum und $U \leq V$ ein fest gewählter Unterraum. Dann ist durch*

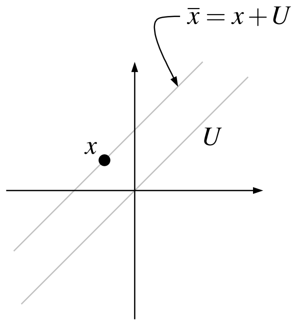
$$x \sim y \quad :\Leftrightarrow \quad x - y \in U \qquad \text{für alle } x, y \in V$$

eine Äquivalenzrelation auf V definiert. Für die Äquivalenzklasse eines Vektors $x \in V$ bezüglich dieser Relation gilt

$$\overline{x} = x + U := \{x + u : u \in U\}.$$

*Man nennt diese Menge (wie im Bild rechts) einen **affinen** bzw. **verschobenen Unterraum** mit Aufpunkt x.*

Die Menge $V/\sim$ aller Äquivalenzklassen bezüglich dieser Relation bezeichnet man mit V/U.



Beweis. Wir zeigen zunächst, dass die gegebene Relation eine Äquivalenzrelation wie in Definition 2.27 ist.

Reflexivität: Für alle $x \in V$ gilt $x - x = 0 \in U$ nach Bemerkung 15.7, und damit $x \sim x$.

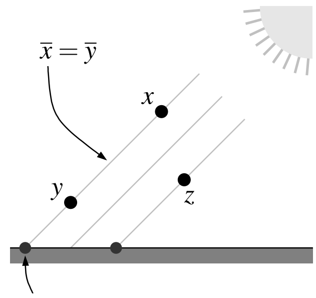
Symmetrie: Es seien $x, y \in V$ mit $x \sim y$, also $x - y \in U$. Dann ist nach Definition 15.5 (a) auch $(-1)(x - y) = y - x \in U$, und damit $y \sim x$.

Transitivität: Sind $x, y, z \in V$ mit $x \sim y$ und $y \sim z$, also $x - y \in U$ und $y - z \in U$, so ist nach Definition 15.5 (a) auch $(x - y) + (y - z) = x - z \in U$, und damit $x \sim z$.

Also ist $\sim$ eine Äquivalenzrelation. Für die Klasse $\overline{x}$ eines Vektors x gilt nun nach Definition 2.27 (c)

$$\overline{x} = \{y \in V : y - x \in U\} = \{y \in V : y - x = u \text{ für ein } u \in U\} = \{x + u : u \in U\}. \qquad \square$$

Bemerkung 17.13 (Anschauliche Deutung von V/U)**.** Die geometrische Bedeutung des Raumes V/U lässt sich am besten wie im Bild rechts erläutern, in dem $V = \mathbb{R}^2$ und $U = \operatorname{Lin}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ist. Dort scheint die Sonne mit parallelen (hell eingezeichneten) Strahlen in Richtung von U und wirft dabei von jedem Punkt in V einen Schatten auf den Boden. In diesem Bild ist die Klasse $\overline{x} \in V/U$ eines Punktes $x \in V$ gerade der Sonnenstrahl durch x. Zwei Punkte in V bestimmen also genau dann den gleichen Punkt in V/U, wenn sie auf dem gleichen Sonnenstrahl liegen, d. h. denselben Schattenpunkt auf dem Boden werfen. Im Bild rechts ist also $\overline{x} = \overline{y} \neq \overline{z}$.



Schattenpunkt von x und y

In diesem Sinne kann man sich V/U damit als eine „Schattenwelt“ von V vorstellen, die zwar jeden Punkt von V sieht, aber nur mit einem Teil seiner Informationen: Der „Abstand zur Sonne“ eines Punktes in V ist anhand des Schattenbildes nicht mehr zu rekonstruieren. Für einen Vektor $x \in V$ nimmt die Klasse $\overline{x}$ also wie beabsichtigt „den Anteil in U heraus“.

Bemerkung 17.14. Für zwei Vektoren $x, y \in V$ gilt nach Satz 2.29 (a) genau dann $\overline{x} = \overline{y}$ in V/U, wenn $x \sim y$ ist. Wir sehen mit Definition 17.12 also für alle $x, y \in V$ in V/U:

$$\boxed{\begin{array}{rccc} & \overline{x} = \overline{y} & \Leftrightarrow & x - y \in U, \\ \text{insbesondere also} & \overline{x} = \overline{0} & \Leftrightarrow & x \in U. \end{array}}$$

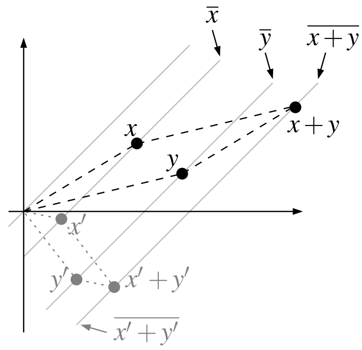
Mit diesen Rechenregeln kann man Gleichungen zwischen Äquivalenzklassen in V/U immer auf Aussagen über die Repräsentanten in V zurückführen. So ist in der Situation von Bemerkung 17.13 beispielsweise

$$\overline{\begin{pmatrix} 3 \\ 5 \end{pmatrix}} = \overline{\begin{pmatrix} 6 \\ 8 \end{pmatrix}}, \quad \text{weil} \quad \begin{pmatrix} 3 \\ 5 \end{pmatrix} - \begin{pmatrix} 6 \\ 8 \end{pmatrix} = \begin{pmatrix} -3 \\ -3 \end{pmatrix} \in U.$$

Allerdings fehlt uns noch ein letzter Schritt: Bisher ist der Raum V/U nur eine Menge ohne weitere Struktur. Um ihn im Rahmen der linearen Algebra untersuchen zu können, müssen wir ihn selbst wieder zu einem Vektorraum machen, also auf ihm eine Vektoraddition und Skalarmultiplikation definieren und zeigen, dass damit dann die Vektorraumeigenschaften für V/U gelten.

Die Idee hierfür ist sehr einfach und im Bild rechts dargestellt: Wollen wir die verschobenen Unterräume $\overline{x}$ und $\overline{y}$ in V/U addieren, so addieren wir einfach wie im oberen Teil des Bildes die Aufpunkte x und y und verwenden den so erhaltenen Punkt $x+y$ als Aufpunkt für die Summe, d. h. wir setzen

$$\overline{x}+\overline{y} := \overline{x+y}. \qquad (*)$$



Allerdings müssen wir dabei etwas aufpassen: Wir hätten für dieselben verschobenen Unterräume statt x und y ja auch wie im unteren Teil des Bildes genauso gut andere Aufpunkte x' bzw. y' wählen können und hätten dann als Ergebnis den verschobenen Unterraum $\overline{x'+y'}$ erhalten!

Damit die Vorschrift $(*)$ wirklich widerspruchsfrei eine Verknüpfung auf V/U definiert, müssen wir also überprüfen, dass der verschobene Unterraum $\overline{x'+y'}$ derselbe ist wie $\overline{x+y}$, d. h. dass *das Endergebnis nicht von der Wahl der Aufpunkte abhängt*. Man sagt dazu auch, dass wir die *Wohldefiniertheit* von $(*)$ überprüfen müssen. Eine solche Überprüfung ist immer dann nötig, wenn wir eine Funktion auf einer Menge von Äquivalenzklassen (hier: V/U) definieren wollen und bei der Konstruktion die Wahl eines Repräsentanten einer Äquivalenzklasse (hier: eines Aufpunkts eines verschobenen Unterraums) verwenden. Die allgemeine Situation ist die folgende:

Notation 17.15 (Wohldefiniertheit)**.** Es sei $\sim$ eine Äquivalenzrelation auf einer Menge M. Möchte man auf der Menge $M/\sim$ der Äquivalenzklassen eine Abbildung in eine andere Menge N definieren, so ist die Idee hierfür in der Regel, dass man eine Abbildung $g\colon M \to N$ wählt und dann

$$f\colon M/\sim \to N,\ f(\overline{x}) := g(x) \qquad (*)$$

setzt. Man möchte das Bild einer Äquivalenzklasse unter f also dadurch definieren, dass man einen Repräsentanten dieser Klasse wählt und diesen dann mit g abbildet. Damit dies nun f widerspruchsfrei definiert, brauchen wir offensichtlich, dass das Ergebnis dieser Vorschrift nicht von der Wahl des Repräsentanten abhängt: Sind $x,y \in M$ äquivalent zueinander, sind sie also Repräsentanten derselben Äquivalenzklasse, so muss $g(x) = g(y)$ gelten. Mit anderen Worten benötigen wir

$$g(x) = g(y) \quad \text{für alle } x,y \in M \text{ mit } \overline{x} = \overline{y},$$

damit die Definition $(*)$ widerspruchsfrei ist. Statt „widerspruchsfrei“ sagt man in diesem Fall wie oben schon erwähnt in der Regel, dass f durch die Vorschrift $(*)$ **wohldefiniert** ist. Die Wohldefiniertheit einer Funktion muss man also immer dann nachprüfen, wenn der Startraum der Funktion eine Menge von Äquivalenzklassen ist und die Funktionsvorschrift Repräsentanten dieser Klassen benutzt. Oder noch etwas allgemeiner: Wenn eine Funktionsvorschrift an irgendeiner Stelle eine Wahl beinhaltet, muss man sich vergewissern, dass der letztliche Funktionswert von dieser Wahl unabhängig ist.

Nach diesen Vorbemerkungen können wir die Menge V/U nun wie angekündigt zu einem Vektorraum machen:

Satz und Definition 17.16 (Quotientenräume)**.** *Es sei U ein Unterraum eines K-Vektorraums V. Dann sind die Verknüpfungen*

$$\overline{x}+\overline{y} := \overline{x+y} \qquad \text{und} \qquad \lambda \cdot \overline{x} := \overline{\lambda x} \qquad \text{für } x,y \in V \text{ und } \lambda \in K$$

*auf V/U wohldefiniert und machen V/U zu einem K-Vektorraum. Man nennt ihn den **Quotientenraum** bzw. **Faktorraum** von V nach U.*

*Man spricht V/U auch als „V **modulo** U“ und nennt $\overline{x} \in V/U$ für ein $x \in V$ die **Restklasse** von V modulo U.*

Beweis. Wir zeigen zunächst die Wohldefiniertheit der Addition: Sind $x, x', y, y' \in V$ mit $\overline{x} = \overline{x'}$ und $\overline{y} = \overline{y'}$, so bedeutet dies nach Bemerkung 17.14 genau $x - x' \in U$ und $y - y' \in U$. Nach Definition 15.5 (a) ist dann aber auch $(x+y) - (x'+y') = (x-x') + (y-y') \in U$ – was wiederum nach Bemerkung 17.14 genau $\overline{x+y} = \overline{x'+y'}$ bedeutet. Also ist die Addition auf V/U wohldefiniert.

Genauso zeigt man die Wohldefiniertheit der Skalarmultiplikation: Sind $\lambda \in K$ und $x, x' \in V$ mit $\overline{x} = \overline{x'}$, also $x - x' \in U$, so ist nach Definition 15.5 (a) auch $\lambda x - \lambda x' = \lambda(x - x') \in U$ und damit $\overline{\lambda x} = \overline{\lambda x'}$.

Die Vektorraumaxiome für V/U ergeben sich nun unmittelbar aus denen von V. So erhält man z. B. die Assoziativität der Vektoraddition durch die einfache Rechnung

$$\left(\overline{x} + \overline{y}\right) + \overline{z} = \overline{x+y} + \overline{z} = \overline{(x+y)+z} = \overline{x+(y+z)} = \overline{x} + \overline{y+z} = \overline{x} + \left(\overline{y} + \overline{z}\right)$$

für alle $x, y, z \in V$, wobei die mittlere Gleichheit die Assoziativität in V ist und sich die anderen Gleichungen aus der Definition der Addition in V/U ergeben. Die übrigen Eigenschaften überprüft man genauso; der Nullvektor in V/U ist die Klasse $\overline{0}$ des Nullvektors in V bzw. der unverschobene Unterraum U, das additive Inverse eines Elements $\overline{x} \in V/U$ ist $\overline{-x}$. □

Aufgrund der anschaulichen Deutung von Komplementen und Quotientenräumen sollte es nicht verwundern, dass wir nun zeigen können, dass diese beiden Konzepte letztlich das gleiche beschreiben, also als Vektorräume isomorph zueinander sind. Wie oben schon erwähnt ist der Vorteil des Komplements lediglich, dass es als Unterraum des ursprünglichen Vektorraums anschaulich leichter zu verstehen ist; der Vorteil des Quotientenraums ist dagegen, dass er ohne willkürliche Wahlen konstruiert werden kann und damit aus mathematischer Sicht das natürlichere Objekt ist.

Satz 17.17 (Dimension von Quotientenräumen)**.** *Es seien U ein Untervektorraum eines K-Vektorraums V und U' ein Komplement von U. Dann ist die Abbildung*

$$f : U' \to V/U, \; x \mapsto \overline{x}$$

ein Isomorphismus.

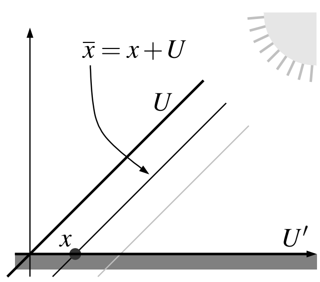
Ist V zusätzlich endlich-dimensional, so gilt also insbesondere $\dim V/U = \dim V - \dim U$.

Beweis. Um zu zeigen, dass f ein Isomorphismus ist, müssen wir die folgenden Dinge überprüfen:

- f ist eine lineare Abbildung, denn für alle $x, y \in U'$ ist
$$f(x+y) = \overline{x+y} = \overline{x} + \overline{y} = f(x) + f(y),$$
und eine analoge Aussage gilt natürlich für die Skalarmultiplikation.
- f ist injektiv: Es sei $x \in U'$ mit $f(x) = \overline{x} = \overline{0}$, also $x \in U$ nach Bemerkung 17.14. Dann ist aber $x \in U \cap U' = \{0\}$. Damit ist f nach Lemma 15.23 injektiv.
- f ist surjektiv: Es sei $\overline{x} \in V/U$ beliebig, also $x \in V$. Wegen $V = U + U'$ können wir $x = x_1 + x_2$ mit $x_1 \in U$ und $x_2 \in U'$ schreiben. Dann liegt x_2 in der Definitionsmenge U' von f, und es gilt $f(x_2) = \overline{x_2} = \overline{x}$ nach Bemerkung 17.14, da $x - x_2 = x_1 \in U$. Also ist f surjektiv.

Die Zusatzaussage folgt nun sofort daraus, dass das Komplement U' nach Bemerkung 17.8 die Dimension $\dim V - \dim U$ hat. □

Bemerkung 17.18. Das Bild rechts illustriert in der Situation von Bemerkung 17.13 noch einmal, dass der Morphismus f aus Satz 17.17 bijektiv ist: Die Bodenlinie U' ist nach Beispiel 17.9 ein Komplement der Richtung U der Sonnenstrahlen. Die Abbildung f ordnet nun jedem Punkt $x \in U'$ auf dem Boden den Sonnenstrahl $\overline{x} \in V/U$ durch diesen Punkt zu, und liefert offensichtlich eine Bijektion zwischen den Bodenpunkten und der Menge der Sonnenstrahlen. Wenn wir in Bemerkung 17.13 gesagt haben, dass V/U die „Schattenwelt“ auf dem Boden ist, haben wir dabei also schon den Isomorphismus zwischen dem eigentlichen Quotientenraum V/U und dem Boden U' verwendet.



Bemerkung 17.19 (Basen von Quotientenräumen)**.** Nach Satz 17.17 (und Lemma 16.15) erhalten wir eine Basis des Quotientenraums V/U, indem wir eine Basis eines Komplements von U wählen und die Restklassen dieser Vektoren in V/U nehmen. Kombinieren wir dies mit dem Verfahren aus Bemerkung 17.11, so bedeutet dies: Wir können eine Basis von V/U konstruieren, indem wir eine Basis von U zu einer Basis von V ergänzen und dann die Restklassen der hinzu genommenen Vektoren wählen. Im Beispiel von Bemerkung 17.11 ist also z. B. $(\overline{x^2})$ eine Basis von V/U.

17.B Isomorphiesätze und Dimensionsformeln

Als Anwendung der Quotientenräume werden wir nun einige nützliche Isomorphismen zwischen unseren bisherigen Konstruktionen herleiten, aus denen dann weitere wichtige Dimensionsformeln folgen. Als Erstes wollen wir dazu den sogenannten Homomorphiesatz beweisen, der aus jeder linearen Abbildung $f\colon V \to W$ „einen Isomorphismus machen kann". Die Idee hierfür ist sehr einfach: Natürlich kann man f zunächst einmal surjektiv machen, indem man den Zielraum W durch den Unterraum $\operatorname{Im} f$ ersetzt. Um f auch noch injektiv zu machen, also gemäß Lemma 15.23 den Kern zu $\{0\}$ zu machen, können wir einfach den Startraum V durch den Quotientenraum $V/\operatorname{Ker} f$ ersetzen: Auf diese Art werden alle Elemente des Kerns von f miteinander identifiziert, so dass der Kern der neuen Abbildung auf dem Quotientenraum nur noch aus dem einen Element $\overline{0} = \operatorname{Ker} f$ besteht.

Satz 17.20 (**Homomorphiesatz**)**.** *Für jede lineare Abbildung $f\colon V \to W$ ist die Abbildung*

$$g\colon V/\operatorname{Ker} f \to \operatorname{Im} f,\ \overline{x} \mapsto f(x)$$

(wohldefiniert und) ein Isomorphismus.

Beweis. Wir müssen einige Dinge überprüfen:

- Die Abbildung g ist wohldefiniert: Sind $x,y \in V$ mit $\overline{x} = \overline{y}$, also $x-y \in \operatorname{Ker} f$ nach Bemerkung 17.14, so ist $f(x-y) = f(x) - f(y) = 0$ und damit $f(x) = f(y)$.
- Die Abbildung g ist linear: Für $x,y \in V$ gilt
$$g(\overline{x}+\overline{y}) = g(\overline{x+y}) = f(x+y) = f(x)+f(y) = g(\overline{x})+g(\overline{y});$$
analog folgt auch die Verträglichkeit mit der Skalarmultiplikation.
- Die Abbildung g ist surjektiv: Dies ist klar nach Definition von $\operatorname{Im} f$, denn jedes Element in $\operatorname{Im} f$ ist ja von der Form $f(x) = g(\overline{x})$ für ein $x \in V$.
- Die Abbildung g ist injektiv: Nach Lemma 15.23 genügt es dafür zu zeigen, dass $\operatorname{Ker} g = \{\overline{0}\}$. Es sei also $x \in V$ mit $g(\overline{x}) = 0$. Dann ist $f(x) = 0$, also $x \in \operatorname{Ker} f$ und damit $\overline{x} = \overline{0} \in V/\operatorname{Ker} f$ nach Bemerkung 17.14. □

Beispiel 17.21 (Anschauliche Deutung des Homomorphiesatzes)**.** Als anschauliches Beispiel für den Homomorphiesatz können wir noch einmal die „Schattenwelt" aus Bemerkung 17.13 und Bemerkung 17.18 betrachten. Ist $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ die lineare Abbildung, die einen Punkt auf seinen Schattenpunkt auf den Boden abbildet, so ist $\operatorname{Ker} f = U$ der Sonnenstrahl durch 0 und $\operatorname{Im} f = U'$ der Boden. Satz 17.20 gibt uns dann den Isomorphismus $g\colon \mathbb{R}^2/U \to U'$, der jeden Sonnenstrahl auf seinen Bodenpunkt abbildet und genau die Umkehrung des Isomorphismus aus Satz 17.17 ist.

Aufgabe 17.22. Die lineare Abbildung, die der Situation in Beispiel 17.21 entspricht, ist

$$f\colon \mathbb{R}^2 \to \mathbb{R}^2,\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 - x_2 \\ 0 \end{pmatrix}.$$

Überprüfe den Homomorphiesatz in diesem Fall explizit, d. h. zeige durch eine direkte Rechnung, dass die Abbildung

$$g\colon \mathbb{R}^2/\operatorname{Lin}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \to \operatorname{Lin}\begin{pmatrix} 1 \\ 0 \end{pmatrix},\ \overline{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}} \mapsto \begin{pmatrix} x_1 - x_2 \\ 0 \end{pmatrix}$$

wohldefiniert, linear, surjektiv und injektiv ist.

Folgerung 17.23 (Dimensionsformel für Morphismen)**.** *Für jeden Morphismus $f\colon V \to W$ zwischen endlich-dimensionalen K-Vektorräumen gilt* $\dim \operatorname{Im} f + \dim \operatorname{Ker} f = \dim V$.

Beweis. Nach dem Homomorphiesatz 17.20 ist $V / \operatorname{Ker} f$ isomorph zu $\operatorname{Im} f$. Also gilt

$$\dim \operatorname{Im} f = \dim(V / \operatorname{Ker} f) = \dim V - \dim \operatorname{Ker} f$$

nach Satz 17.17. □

Nach Kernen und Bildern linearer Abbildungen betrachten wir als Nächstes Summen und Durchschnitte von Unterräumen. Auch dafür gibt es zunächst wieder eine nützliche Isomorphieaussage.

Satz 17.24 (**Isomorphiesatz** für Summen und Durchschnitte)**.** *Es seien U_1 und U_2 zwei Unterräume eines K-Vektorraums V. Dann ist die Abbildung*

$$g\colon U_1/(U_1 \cap U_2) \to (U_1 + U_2)/U_2,\ \overline{x} \mapsto \overline{x}$$

(wohldefiniert und) ein Isomorphismus.

Beweis. Wir betrachten zunächst die Abbildung

$$f\colon U_1 \to (U_1 + U_2)/U_2,\ x \mapsto \overline{x}.$$

Sie ist offensichtlich linear (und Wohldefiniertheit muss hier nicht überprüft werden, da der Startraum kein Quotientenraum ist). Wir bestimmen das Bild und den Kern von f:

- $\operatorname{Im} f = (U_1 + U_2)/U_2$, d. h. f ist surjektiv: Für ein beliebiges Element $\overline{x_1 + x_2} \in (U_1 + U_2)/U_2$ (mit $x_1 \in U_1$ und $x_2 \in U_2$) ist $x_1 \in U_1$ ein Urbild unter f, denn $f(x_1) = \overline{x_1} = \overline{x_1 + x_2}$.
- $\operatorname{Ker} f = U_1 \cap U_2$: Ein Element $x \in U_1$ liegt genau dann im Kern von f, wenn $f(x) = \overline{x} = \overline{0}$ ist, was nach Bemerkung 17.14 äquivalent ist zu $x \in U_2$. Der Kern von f besteht also aus allen Elementen von U_1, die auch in U_2 liegen, und ist damit gleich $U_1 \cap U_2$.

Der Homomorphiesatz 17.20 für f liefert nun den Isomorphismus $g\colon U_1 / \operatorname{Ker} f \to \operatorname{Im} f,\ \overline{x} \mapsto f(x)$, also wie behauptet

$$g\colon U_1/(U_1 \cap U_2) \to (U_1 + U_2)/U_2,\ \overline{x} \mapsto \overline{x}.$$ □

Bemerkung 17.25. In der Literatur sind für die Sätze 17.20 und 17.24 z. T. recht unterschiedliche Bezeichnungen üblich. Manche Bücher bezeichnen sie als den 1. und 2. Isomorphiesatz, während andere aber auch Satz 17.24 den 1. Isomorphiesatz nennen.

Folgerung 17.26 (Dimensionsformel für Summen und Durchschnitte)**.** *Sind U_1 und U_2 endlich erzeugte Unterräume eines K-Vektorraums V, so gilt*

$$\dim(U_1 + U_2) + \dim(U_1 \cap U_2) = \dim U_1 + \dim U_2.$$

Beweis. Nach dem Isomorphiesatz 17.24 haben $U_1/(U_1 \cap U_2)$ und $(U_1 + U_2)/U_2$ dieselbe Dimension. Aus der Dimensionsformel für Quotienten in Satz 17.17 ergibt sich also

$$\dim U_1 - \dim(U_1 \cap U_2) = \dim(U_1 + U_2) - \dim U_2$$

und somit die Behauptung. □

Beispiel 17.27. Beachte, dass nur die Summe der Dimensionen von $U_1 \cap U_2$ und $U_1 + U_2$ durch $\dim U_1$ und $\dim U_2$ bestimmt sind, nicht aber die Dimensionen von $U_1 \cap U_2$ und $U_1 + U_2$ selbst. Als einfaches Beispiel hierfür seien U_1 und U_2 zwei Geraden (durch den Ursprung) in $\mathbb{R}^2$. Dann gibt es zwei Möglichkeiten:

(a) Ist $U_1 = U_2$, so ist $U_1 + U_2 = U_1 \cap U_2 = U_1 = U_2$, und die Dimensionsformel ergibt

$$\dim(U_1 + U_2) + \dim(U_1 \cap U_2) = 1 + 1 = 1 + 1 = \dim U_1 + \dim U_2.$$

(b) Ist hingegen $U_1 \neq U_2$, so ist $U_1 + U_2 = \mathbb{R}^2$ und $U_1 \cap U_2 = \{0\}$, und damit

$$\dim(U_1 + U_2) + \dim(U_1 \cap U_2) = 2 + 0 = 1 + 1 = \dim U_1 + \dim U_2.$$

Aufgabe 17.28. Es seien V ein K-Vektorraum und $U_1, U_2 \leq V$ mit $\dim V = 6$, $\dim U_1 = 5$ und $\dim U_2 = 3$.

Welche Dimension kann $U_1 \cap U_2$ haben? Gib für jede solche Möglichkeit ein konkretes Beispiel für U_1, U_2 und V an.

18. Determinanten

Wir haben nun schon viele Konzepte der linearen Algebra untersucht und gesehen, wie man – im endlich-dimensionalen Fall in der Regel mit Hilfe von Matrizen – mit ihnen konkret rechnen kann. Daher wollen wir jetzt zum Abschluss des ersten Teils der linearen Algebra noch die sogenannten Determinanten einführen, die beim Rechnen mit Matrizen ein unverzichtbares Hilfsmittel sind. Determinanten haben sehr viele schöne Eigenschaften und können demzufolge auch auf viele verschiedene Arten motiviert werden. Eine mögliche Herangehensweise ist, dass man nach einem einfachen Kriterium für die Invertierbarkeit quadratischer Matrizen sucht, so wie in dem folgenden einfachen Lemma für 2×2-Matrizen:

Lemma 18.1. *Eine* 2×2*-Matrix*

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}$$

über einem Körper K *ist genau dann invertierbar, wenn* $a_{1,1}a_{2,2} - a_{1,2}a_{2,1} \neq 0$.

Beweis. Wir unterscheiden zwei Fälle:

1. Fall: Ist $a_{1,1} = 0$, so ist A genau dann invertierbar, wenn $a_{1,2} \neq 0$ und $a_{2,1} \neq 0$ gilt – denn wenn diese beiden Einträge ungleich Null sind, sind die beiden Spalten von A offensichtlich linear unabhängig (so dass dann $\operatorname{rk} A = 2$ gilt), während A andernfalls eine Nullzeile oder Nullspalte enthält und damit höchstens Rang 1 haben kann. Nach Lemma 14.11 ist A also genau dann invertierbar, wenn $a_{1,2} \neq 0$ und $a_{2,1} \neq 0$, also $a_{1,1}a_{2,2} - a_{1,2}a_{2,1} \neq 0$ gilt.

2. Fall: Ist $a_{1,1} \neq 0$, so wenden wir den Gauß-Algorithmus an, um $\operatorname{rk} A$ zu berechnen:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \xrightarrow{Z1 \to \frac{1}{a_{1,1}} Z1} \begin{pmatrix} 1 & \frac{a_{1,2}}{a_{1,1}} \\ a_{2,1} & a_{2,2} \end{pmatrix} \xrightarrow{Z2 \to Z2 - a_{2,1} Z1} \begin{pmatrix} 1 & \frac{a_{1,2}}{a_{1,1}} \\ 0 & a_{2,2} - \frac{a_{1,2}a_{2,1}}{a_{1,1}} \end{pmatrix}.$$

Diese Matrix hat genau dann Rang 2, ist also genau dann invertierbar, wenn $a_{2,2} - \frac{a_{1,2}a_{2,1}}{a_{1,1}} \neq 0$, d. h. wenn $a_{1,1}a_{2,2} - a_{1,2}a_{2,1} \neq 0$ gilt. □

Die Zahl $a_{1,1}a_{2,2} - a_{1,2}a_{2,1}$ werden wir später die *Determinante* $\det A$ von A nennen (weil sie „determiniert“, ob A invertierbar ist oder nicht).

Unser Ziel in diesem Kapitel ist es, Lemma 18.1 auf größere quadratische Matrizen zu verallgemeinern, also zu jeder Matrix $A \in K^{n \times n}$ eine Zahl $\det A \in K$ zu definieren, die ein Polynom in den Einträgen von A ist und (neben vielen anderen schönen Eigenschaften) genau dann ungleich Null ist, wenn A invertierbar ist.

18.A Konstruktion der Determinante

Leider ist eine direkte Angabe der Determinante einer quadratischen Matrix $A \in K^{n \times n}$ als polynomialer Ausdruck in den Einträgen von A so wie in Lemma 18.1 für allgemeines n zwar möglich (siehe Bemerkung 18.14), aber auch recht kompliziert. Wir wollen daher hier den für euch wahrscheinlich etwas ungewohnten Zugang wählen, die Determinante über ihre Eigenschaften zu definieren, d. h. als eine Funktion $A \mapsto \det A$ auf den $n \times n$-Matrizen, die eine gewisse „Wunschliste“ elementarer Eigenschaften erfüllt. Im Anschluss werden wir dann zeigen, dass unsere Wunschliste wirklich erfüllbar ist und die Determinante in der Tat auch eindeutig bestimmt.

Hier ist nun unsere Wunschliste:

Definition 18.2 (Determinante)**.** Es seien K ein Körper und $n \in \mathbb{N}_{>0}$ gegeben. Eine Abbildung $\det\colon K^{n \times n} \to K$ heißt **Determinante** (von $n \times n$-Matrizen), wenn gilt:

(a) („det ist multilinear“) Die Funktion det ist *linear in jeder Zeile*, d. h. für alle $k \in \{1,\dots,n\}$ und $\lambda \in K$ gilt

$$\det \begin{pmatrix} a_1 \\ \vdots \\ a_k + a'_k \\ \vdots \\ a_n \end{pmatrix} = \det \begin{pmatrix} a_1 \\ \vdots \\ a_k \\ \vdots \\ a_n \end{pmatrix} + \det \begin{pmatrix} a_1 \\ \vdots \\ a'_k \\ \vdots \\ a_n \end{pmatrix} \quad \text{und} \quad \det \begin{pmatrix} a_1 \\ \vdots \\ \lambda\, a_k \\ \vdots \\ a_n \end{pmatrix} = \lambda \cdot \det \begin{pmatrix} a_1 \\ \vdots \\ a_k \\ \vdots \\ a_n \end{pmatrix},$$

wobei $a_1,\dots,a_k,a'_k,\dots,a_n \in K^{1\times n}$ die Zeilen der jeweiligen (quadratischen) Matrizen bezeichnen. (Halten wir also alle Zeilen bis auf die k-te fest, so haben wir genau eine lineare Abbildung in der k-ten Zeile im Sinne von Definition 15.16.)

(b) („det ist alternierend“) Stimmen zwei Zeilen von $A \in K^{n\times n}$ überein, so ist $\det A = 0$.

(c) („det ist normiert“) Es gilt $\det(E_n) = 1$.

Beispiel 18.3. Die Funktion

$$\det\colon K^{2\times 2} \to K,\ \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \mapsto a_{1,1}a_{2,2} - a_{1,2}a_{2,1}$$

aus Lemma 18.1 ist eine Determinante:

(a) det ist multilinear: Die Additivität in der ersten Zeile ergibt sich z. B. aus der Rechnung

$$\begin{aligned} \det \begin{pmatrix} a_{1,1}+a'_{1,1} & a_{1,2}+a'_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} &= (a_{1,1}+a'_{1,1})\,a_{2,2} - (a_{1,2}+a'_{1,2})\,a_{2,1} \\ &= a_{1,1}a_{2,2} - a_{1,2}a_{2,1} + a'_{1,1}a_{2,2} - a'_{1,2}a_{2,1} \\ &= \det \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} + \det \begin{pmatrix} a'_{1,1} & a'_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}; \end{aligned}$$

die anderen Linearitätseigenschaften folgen natürlich genauso.

(b) det ist alternierend: Sind die beiden Zeilen der Matrix gleich, so ist

$$\det \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{1,1} & a_{1,2} \end{pmatrix} = a_{1,1}a_{1,2} - a_{1,2}a_{1,1} = 0.$$

(c) det ist normiert, denn natürlich ist $\det(E_2) = 1$.

Bemerkung 18.4.

(a) Wir werden in Folgerung 18.8 und Satz 18.12 sehen, dass es zu jedem Körper K und jedem $n \in \mathbb{N}_{>0}$ in der Tat genau eine Determinante $\det\colon K^{n\times n} \to K$ gibt, dass Definition 18.2 die Determinante also widerspruchsfrei und eindeutig festlegt. Solange wir dies noch nicht gezeigt haben, sollten wir aber korrekterweise immer von *einer* Determinante (und nicht von *der* Determinante) sprechen.

(b) Enthält A eine Nullzeile, so können wir aus dieser Zeile den Faktor 0 herausziehen und erhalten aus der Linearitätseigenschaft in dieser Zeile sofort, dass dann $\det A = 0$ sein muss.

(c) Aus Eigenschaft (b) der Definition 18.2 einer Determinante folgt, dass sich $\det A$ beim Vertauschen zweier Zeilen mit -1 multipliziert, also genau das Vorzeichen ändert (daher kommt auch der Name „alternierend“ für diese Eigenschaft): Für alle $k,l \in \{1,\dots,n\}$ mit $k \neq l$ ergibt sich zusammen mit der Multilinearität nämlich

$$\underbrace{\det\begin{pmatrix}\vdots\\a_k+a_l\\\vdots\\a_k+a_l\\\vdots\end{pmatrix}}_{=0}=\det\begin{pmatrix}\vdots\\a_k\\\vdots\\a_k+a_l\\\vdots\end{pmatrix}+\det\begin{pmatrix}\vdots\\a_l\\\vdots\\a_k+a_l\\\vdots\end{pmatrix}$$

$$=\underbrace{\det\begin{pmatrix}\vdots\\a_k\\\vdots\\a_k\\\vdots\end{pmatrix}}_{=0}+\det\begin{pmatrix}\vdots\\a_k\\\vdots\\a_l\\\vdots\end{pmatrix}+\det\begin{pmatrix}\vdots\\a_l\\\vdots\\a_k\\\vdots\end{pmatrix}+\underbrace{\det\begin{pmatrix}\vdots\\a_l\\\vdots\\a_l\\\vdots\end{pmatrix}}_{=0},$$

und damit

$$\det\begin{pmatrix}\vdots\\a_k\\\vdots\\a_l\\\vdots\end{pmatrix}=-\det\begin{pmatrix}\vdots\\a_l\\\vdots\\a_k\\\vdots\end{pmatrix}.$$

(d) Analog zu (c) wollen wir jetzt untersuchen, was mit einer Determinante passiert, wenn wir in einer Matrix A für gegebenes $k \in \{1,\dots,n\}$ die k-te Zeile unter Beibehaltung der Reihenfolge der anderen Zeilen ganz nach oben schieben. Wir können dies wie folgt durch $k-1$ Vertauschungen zweier benachbarter Zeilen erreichen:

$$A=\begin{pmatrix}\vdots\\a_{k-2}\\a_{k-1}\\a_k\\a_{k+1}\\\vdots\end{pmatrix}\longrightarrow\begin{pmatrix}\vdots\\a_{k-2}\\a_k\\a_{k-1}\\a_{k+1}\\\vdots\end{pmatrix}\longrightarrow\cdots\longrightarrow\begin{pmatrix}a_k\\a_1\\\vdots\\a_{k-1}\\a_{k+1}\\\vdots\end{pmatrix}.$$

Da sich bei jeder dieser Vertauschungen nach (c) das Vorzeichen der Determinante ändert, ändert das gesamte Verschieben der k-ten Zeile ganz nach oben die Determinante von A also um einen Faktor $(-1)^{k-1}$.

Um die weiteren Eigenschaften von Determinanten zu untersuchen, beginnen wir zunächst mit den Elementarmatrizen.

Lemma 18.5 (Determinanten von Elementarmatrizen)**.** *Es sei* $\det\colon K^{n\times n}\to K$ *eine Determinante. Dann gilt für alle* $A\in K^{n\times n}$ *sowie für alle* $n\times n$*-Elementarmatrizen* $F_k(\lambda)$ *und* $F_{k,l}(\lambda)$ *aus Konstruktion 14.1:*

(a) $\det(F_k(\lambda)\cdot A)=\lambda\,\det A$.

(b) $\det(F_{k,l}(\lambda)\cdot A)=\det A$.

Insbesondere gilt also $\det F_k(\lambda)=\lambda$ *und* $\det F_{k,l}(\lambda)=1$*, und damit* $\det(FA)=\det F\cdot\det A$ *für jede Elementarmatrix* F *und jede beliebige quadratische Matrix* A.

Beweis. Es seien $a_1,\dots,a_n\in K^{1\times n}$ die Zeilen von A. Nach Konstruktion 14.1 entspricht eine Multiplikation von A mit einer Elementarmatrix von links genau einer elementaren Zeilenumformung.

Damit erhalten wir mit den Eigenschaften (a) und (b) aus Definition 18.2

$$\det(F_k(\lambda)\cdot A) = \det\begin{pmatrix} \vdots \\ \lambda\, a_k \\ \vdots \end{pmatrix} = \lambda \det\begin{pmatrix} \vdots \\ a_k \\ \vdots \end{pmatrix} = \lambda \det A$$

und

$$\det(F_{k,l}(\lambda)\cdot A) = \det\begin{pmatrix} \vdots \\ a_k+\lambda\, a_l \\ \vdots \\ a_l \\ \vdots \end{pmatrix} = \det\begin{pmatrix} \vdots \\ a_k \\ \vdots \\ a_l \\ \vdots \end{pmatrix} + \lambda \underbrace{\det\begin{pmatrix} \vdots \\ a_l \\ \vdots \\ a_l \\ \vdots \end{pmatrix}}_{=0} = \det\begin{pmatrix} \vdots \\ a_k \\ \vdots \\ a_l \\ \vdots \end{pmatrix} = \det A,$$

was die beiden Teile des Lemmas zeigt. Die Determinanten der Elementarmatrizen erhält man daraus für $A = E_n$. □

Aus diesem einfachen Lemma folgt nun bereits die wahrscheinlich wichtigste Eigenschaft von Determinanten:

Satz 18.6 (**Produktsatz** für Determinanten)**.** *Es sei* $\det\colon K^{n\times n} \to K$ *eine Determinante. Dann gilt für alle* $A,B \in K^{n\times n}$*:*

(a) $\det(AB) = \det A \cdot \det B$.

(b) *A ist genau dann invertierbar, wenn* $\det A \neq 0$. *In diesem Fall ist* $\det(A^{-1}) = \frac{1}{\det A}$.

Beweis. Wir unterscheiden zwei Fälle:

1. Fall: A ist invertierbar. Dann ist $A = F_1 \cdot \dots \cdot F_k$ nach Lemma 14.11 ein Produkt von Elementarmatrizen. Durch k-fache Anwendung von Lemma 18.5 erhält man dann

$$\begin{aligned} \det(AB) &= \det(F_1\cdot\dots\cdot F_k\cdot B) = \det F_1\cdot\dots\cdot\det F_k\cdot\det B \\ \text{sowie}\qquad \det A &= \det(F_1\cdot\dots\cdot F_k) \qquad = \det F_1\cdot\dots\cdot\det F_k, \end{aligned}$$

und damit wie behauptet $\det(AB) = \det A \cdot \det B$. Setzt man hier $B = A^{-1}$ ein, so ergibt sich insbesondere $\det A \cdot \det A^{-1} = \det(AA^{-1}) = \det E_n = 1$, d. h. es ist $\det A \neq 0$ und $\det(A^{-1}) = \frac{1}{\det A}$.

2. Fall: A ist nicht invertierbar, also $\operatorname{rk} A < n$. Bringen wir A dann mit einem Produkt F von Elementarmatrizen auf Zeilenstufenform FA, so hat FA weniger als n Stufen und damit am Ende (mindestens) eine Nullzeile. Also ist $\det(FA) = 0$ nach Bemerkung 18.4 (b). Da F als Produkt von Elementarmatrizen nach Lemma 14.11 invertierbar ist, bedeutet dies nach dem bereits gezeigten 1. Fall auch $\det F \cdot \det A = 0$; wegen $\det F \neq 0$ also $\det A = 0$.

Mit FA hat aber auch FAB eine Nullzeile. Damit folgt genauso wie oben auch $\det(AB) = 0$, also insbesondere $\det(AB) = \det A \cdot \det B$. □

Bemerkung 18.7. Im Gegensatz zu Produkten gibt es *keine* Formel für die Determinante $\det(A+B)$ einer Summe von zwei Matrizen – insbesondere ist im Allgemeinen $\det(A+B) \neq \det A + \det B$!

Als Folgerung aus dem Produktsatz können wir nun bereits beweisen, dass die Eigenschaften aus Definition 18.2 eine Determinante eindeutig festlegen.

Folgerung 18.8 (Eindeutigkeit der Determinante)**.** *Zu jedem Körper K und $n \in \mathbb{N}_{>0}$ gibt es höchstens eine Determinante* $\det\colon K^{n\times n} \to K$.

Beweis. Es sei $A \in K^{n\times n}$. Ist A nicht invertierbar, so ist nach Satz 18.6 notwendigerweise $\det A = 0$. Andernfalls ist $A = F_1 \cdot \dots \cdot F_k$ nach Lemma 14.11 ein Produkt von Elementarmatrizen, und damit ist nach Satz 18.6 (a)

$$\det A = \det F_1 \cdot \dots \cdot \det F_k.$$

Da die Determinante der Elementarmatrizen nach Lemma 18.5 aber durch Definition 18.2 eindeutig bestimmt ist, ist damit auch $\det A$ durch diese Definition eindeutig festgelegt. □

Auf ganz ähnliche Art wollen wir nun zeigen, dass sich eine Determinante beim Transponieren der Matrizen nicht ändert.

Folgerung 18.9. *Ist $A \in K^{n\times n}$ und $\det\colon K^{n\times n} \to K$ eine Determinante, so gilt $\det(A^\mathsf{T}) = \det A$.*

Beweis. Ist A nicht invertierbar, also $\operatorname{rk} A < n$, so ist nach Folgerung 14.17 auch A^T nicht invertierbar, und damit ist $\det(A^\mathsf{T}) = 0 = \det A$ nach Satz 18.6 (b).

Andernfalls ist $A = F_1 \cdot \dots \cdot F_k$ nach Lemma 14.11 wieder ein Produkt von Elementarmatrizen. Da die zu zeigende Aussage für Elementarmatrizen aus Lemma 18.5 offensichtlich ist (es ist nämlich $(F_k(\lambda))^\mathsf{T} = F_k(\lambda)$ und $(F_{k,l}(\lambda))^\mathsf{T} = F_{l,k}(\lambda)$), folgt somit nach Lemma 13.11 (d) und Satz 18.6 (a)

$$\det(A^\mathsf{T}) = \det((F_1 \cdots F_k)^\mathsf{T}) = \det(F_k^\mathsf{T} \cdots F_1^\mathsf{T}) = \det(F_k^\mathsf{T}) \cdots \det(F_1^\mathsf{T}) = \det F_1 \cdots \det F_k = \det A. \qquad \square$$

Bemerkung 18.10. Folgerung 18.9 besagt anschaulich, dass alle Eigenschaften, die für die Zeilen einer Determinante gelten, analog auch für die Spalten gelten. So ist eine Determinante z. B. auch linear in jeder Spalte (vgl. Definition 18.2 (a)) und ändert ihr Vorzeichen beim Vertauschen zweier Spalten (vgl. Bemerkung 18.4 (c)).

Um sicherzustellen, dass wir mit Definition 18.2 keine in sich widersprüchliche Wunschliste aufgeschrieben haben, kommen wir nun aber endlich zum bereits angekündigten Resultat, dass eine Determinante mit den geforderten Eigenschaften auch wirklich existiert. Wir werden die Funktionen $\det\colon K^{n\times n} \to K$ rekursiv über n definieren und verwenden dazu die folgende Konstruktion, um Matrizen der Größe n auf solche der Größe $n-1$ zurückzuführen.

Definition 18.11 (Streichungsmatrix)**.** Zu $A = (a_{i,j})_{i,j} \in K^{n\times n}$ sowie $k,l \in \{1,\dots,n\}$ sei

$$A'_{k,l} := \begin{pmatrix} a_{1,1} & \cdots & a_{1,l-1} & a_{1,l+1} & \cdots & a_{1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{k-1,1} & \cdots & a_{k-1,l-1} & a_{k-1,l+1} & \cdots & a_{k-1,n} \\ a_{k+1,1} & \cdots & a_{k+1,l-1} & a_{k+1,l+1} & \cdots & a_{k+1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,l-1} & a_{n,l+1} & \cdots & a_{n,n} \end{pmatrix} \in K^{(n-1)\times(n-1)}$$

die Matrix, die man erhält, wenn man aus A die k-te Zeile und l-te Spalte herausstreicht. Wir bezeichnen diese Matrizen als **Streichungsmatrizen** zu A.

Satz 18.12 (Existenz der Determinante)**.** *Für alle $n \in \mathbb{N}_{>0}$ definieren wir $\det\colon K^{n\times n} \to K$ rekursiv über n durch die folgende Vorschrift:*

- *Für $n = 1$ setzen wir $\det(a_{1,1}) := a_{1,1}$.*
- *Für $n > 1$ setzen wir*

$$\det A := \sum_{k=1}^{n} (-1)^{k+1} a_{k,1} \det A'_{k,1},$$

wobei wie üblich $a_{k,1}$ die Einträge der ersten Spalte von A und $A'_{k,1}$ die zu diesen Einträgen gehörigen Streichungsmatrizen sind.

Dann ist $\det$ eine (und damit nach Folgerung 18.8 „die“) Determinante für alle n.

Bevor wir diesen Satz beweisen, wollen wir uns ein paar Beispiele anschauen, um die angegebene rekursive Formel besser zu verstehen.

Beispiel 18.13 (Determinante von 2×2- und 3×3-Matrizen)**.**

(a) Für $n = 2$ besagt die Formel aus Satz 18.12

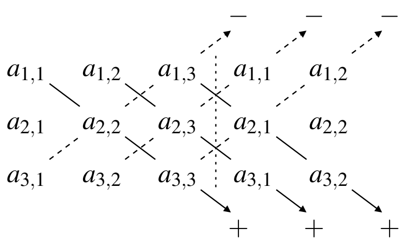
$$\begin{aligned}\det \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} &= (-1)^{1+1} a_{1,1} \det(a_{2,2}) + (-1)^{2+1} a_{2,1} \det(a_{1,2}) \\ &= a_{1,1}a_{2,2} - a_{2,1}a_{1,2}\end{aligned}$$

und reproduziert damit die Formel aus Lemma 18.1.

(b) Für $n = 3$ ergibt sich unter Benutzung des Ergebnisses aus (a)

$$\begin{aligned}\det \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix} &= (-1)^{1+1} a_{1,1} \det \begin{pmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{pmatrix} + (-1)^{2+1} a_{2,1} \det \begin{pmatrix} a_{1,2} & a_{1,3} \\ a_{3,2} & a_{3,3} \end{pmatrix} \\ &\quad + (-1)^{3+1} a_{3,1} \det \begin{pmatrix} a_{1,2} & a_{1,3} \\ a_{2,2} & a_{2,3} \end{pmatrix} \\ &= a_{1,1}a_{2,2}a_{3,3} - a_{1,1}a_{2,3}a_{3,2} - a_{2,1}a_{1,2}a_{3,3} + a_{2,1}a_{1,3}a_{3,2} \\ &\quad + a_{3,1}a_{1,2}a_{2,3} - a_{3,1}a_{1,3}a_{2,2}.\end{aligned}$$

Am einfachsten kann man sich diese Formel nach der sogenannten **Regel von Sarrus** merken: Bilden wir die 3×5-Matrix, in der wir neben der Matrix A die beiden ersten Spalten noch einmal wiederholen, so ergeben sich die 6 Terme der Determinante mit ihren Vorzeichen aus dem folgenden Schema:



Beachte aber, dass diese einfache Merkregel *nur für* $n = 3$ gilt – für größere n ist der komplett ausmultiplizierte Ausdruck für $\det A$ deutlich komplizierter (und für konkrete numerische Berechnungen in der Tat auch nicht mehr geeignet).

Bemerkung 18.14. Diejenigen von euch, die aus der Parallelvorlesung „Algebraische Strukturen“ die symmetrische Gruppe S_n aller Permutationen von $\{1, \dots, n\}$ kennen [G, Kapitel 2], können die Formel für die Determinante auch nicht-rekursiv als

$$\det A = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \cdot a_{1,\sigma(1)} \cdot \dots \cdot a_{n,\sigma(n)} \qquad (*)$$

hinschreiben. Man sieht an dieser Darstellung also, dass die Determinante aus einer Summe von $n!$ Termen besteht. Dabei ist jeder Term ein Produkt von genau n Einträgen von A, und zwar aus jeder Zeile und jeder Spalte genau einem. Aufsummiert wird über alle Möglichkeiten, n Einträge von A eben gerade so auszuwählen, dass man aus jeder Zeile und Spalte einen Eintrag genommen hat. Die Vorzeichen der einzelnen Terme sind immer genau das Vorzeichen der entsprechenden Permutation.

Wir werden die Formel $(*)$ in dieser Vorlesung aber nicht benötigen und daher auch nicht beweisen, dass sie wirklich mit der rekursiven Definition aus Satz 18.12 übereinstimmt bzw. die Eigenschaften von Definition 18.2 erfüllt.

Wir kommen nun aber endlich zum Beweis des Existenzsatzes 18.12.

Beweis von Satz 18.12. Wir überprüfen die drei Eigenschaften aus Definition 18.2 mit Induktion über n. Für $n = 1$ sind alle Aussagen klar. Wir können also annehmen, dass $n > 1$ ist und wir die Eigenschaften von Definition 18.2 für Matrizen der Größe $n - 1$ bereits gezeigt haben; wir müssen sie nun für Matrizen der Größe n zeigen.

det ist multilinear: Der Ausdruck $a_{1,1} \det A'_{1,1}$ ist linear in der ersten Zeile, da $a_{1,1}$ natürlich linear in der ersten Zeile ist und $A'_{1,1}$ nicht von der ersten Zeile abhängt. Die Ausdrücke $a_{k,1} \det A'_{k,1}$ für $k > 1$ sind ebenfalls linear in der ersten Zeile, da $a_{k,1}$ nicht von der ersten Zeile abhängt und $\det A'_{k,1}$ nach Induktionsvoraussetzung linear in der ersten Zeile ist. Damit ist auch $\det A$ als Linearkombination dieser Ausdrücke linear in der ersten Zeile. Die Linearität in den anderen Zeilen folgt natürlich analog.

det ist alternierend: Wir bezeichnen die Zeilen von A mit $a_1, \dots, a_n \in K^{1\times n}$. Weiterhin seien $a'_1, \dots, a'_n \in K^{1\times(n-1)}$ die Zeilen von A, bei denen man jeweils den ersten Eintrag herausgestrichen hat. Wir nehmen nun an, dass zwei Zeilen a_i und a_j von A übereinstimmen, und müssen zeigen, dass $\det A = 0$ folgt. Ohne Beschränkung der Allgemeinheit sei dazu $i > j$.

Beachte, dass dann auch in den Streichungsmatrizen $A'_{k,1}$ mit $k \neq i$ und $k \neq j$, bei denen wir also weder die i-te noch die j-te Zeile herausgestrichen haben, jeweils zwei Zeilen übereinstimmen. Nach Induktionsvoraussetzung ist die Determinante aller dieser Streichungsmatrizen gleich 0, und damit bleibt in der rekursiven Formel für $\det A$ nur der Ausdruck

$$\det A = (-1)^{i+1} a_{i,1} \det A'_{i,1} + (-1)^{j+1} a_{j,1} \det A'_{j,1} \qquad (*)$$

übrig. Nun können wir

$$\text{sowohl} \quad A'_{i,1} = \begin{pmatrix} a'_1 \\ \vdots \\ a'_{j-1} \\ a'_j \\ a'_{j+1} \\ \vdots \\ a'_{i-1} \\ a'_{i+1} \\ \vdots \end{pmatrix} \quad \text{als auch} \quad A'_{j,1} = \begin{pmatrix} a'_1 \\ \vdots \\ a'_{j-1} \\ a'_{j+1} \\ \vdots \\ a'_{i-1} \\ a'_i \\ a'_{i+1} \\ \vdots \end{pmatrix} \quad \text{auf die Form} \quad A' := \begin{pmatrix} a'_i \\ a'_1 \\ \vdots \\ a'_{j-1} \\ a'_{j+1} \\ \vdots \\ a'_{i-1} \\ a'_{i+1} \\ \vdots \end{pmatrix}$$

bringen, indem wir die Zeile a'_j bzw. a'_i unter Beibehaltung der Reihenfolge der anderen Zeilen ganz nach oben schieben. Da det für Matrizen der Größe $n-1$ nach Induktionsvoraussetzung eine Determinante ist, ändern sich dadurch die Vorzeichen von $\det A'_{i,1}$ und $\det A'_{j,1}$ wie in Bemerkung 18.4 (d): Da wir in $A'_{i,1}$ die Zeile mit der Nummer j, in $A'_{j,1}$ jedoch die Zeile mit der Nummer $i-1$ nach oben schieben (im letzteren Fall fehlt ja die Zeile a'_j oberhalb von a'_i), ist also

$$\det A'_{i,1} = (-1)^{j-1} \det A' \quad \text{und} \quad \det A'_{j,1} = (-1)^{i-2} \det A'$$

und damit nach $(*)$

$$\det A = (-1)^{i+j} a_{i,1} \det A' + (-1)^{i+j-1} a_{j,1} \det A' = 0$$

wegen $a_{i,1} = a_{j,1}$.

det ist normiert: In der ersten Spalte der Einheitsmatrix sind natürlich der erste Eintrag gleich 1 und alle anderen gleich 0. Weiterhin ist die Streichungsmatrix des Eintrags links oben gerade E_{n-1}. Also folgt sofort

$$\det E_n = (-1)^{1+1} \cdot 1 \cdot \det E_{n-1} = 1.$$

Damit ist alles gezeigt. □

Insgesamt haben wir jetzt also gesehen, dass es für alle Körper K und $n \in \mathbb{N}$ genau eine Determinante $\det\colon K^{n\times n} \to K$ gibt. In Zukunft werden wir daher immer von *der* Determinante quadratischer Matrizen sprechen.

18.B Eigenschaften der Determinante

Im letzten Abschnitt haben wir die Determinante quadratischer Matrizen definiert und auch bereits ihre ersten wichtigen Eigenschaften gesehen. Wir wollen diese Untersuchung der Determinante jetzt fortsetzen und uns dabei als Erstes um ihre praktische Berechnung kümmern. In der Tat ist hierfür die rekursive Formel aus Satz 18.12 bereits sehr nützlich. Wir können sie allerdings noch etwas erweitern, denn dort ist ja momentan die erste Spalte der Matrix ausgezeichnet – obwohl aufgrund von Definition 18.2 natürlich klar sein sollte, dass die erste Spalte der Matrix keine besondere Rolle spielt. Wir sollten eine ähnliche Rekursionsformel also auch für die anderen Spalten (und aufgrund von Folgerung 18.9 in der Tat auch für die Zeilen) erwarten können. Dies besagt der folgende Satz.

Satz 18.15 (**Laplacescher Entwicklungssatz**)**.** *Es sei $A \in K^{n\times n}$.*

(a) *Für alle $l \in \{1,\dots,n\}$ gilt $\det A = \sum_{k=1}^{n}(-1)^{k+l}\cdot a_{k,l}\cdot \det A'_{k,l}$.*

(b) *Für alle $k \in \{1,\dots,n\}$ gilt $\det A = \sum_{l=1}^{n}(-1)^{k+l}\cdot a_{k,l}\cdot \det A'_{k,l}$.*

Benutzt man diese Formeln, so sagt man auch, dass man die Determinante von A nach der l-ten Spalte bzw. k-ten Zeile entwickelt.

Beweis.

(a) Es sei $B = (b_{i,j})_{i,j}$ die Matrix, die man aus A erhält, indem man die Spalte l unter Beibehaltung der Reihenfolge der anderen Spalten ganz nach links schiebt. Nach den Bemerkungen 18.4 (d) und 18.10 ist dann $\det A = (-1)^{l-1}\det B$. Andererseits ist natürlich $b_{k,1} = a_{k,l}$ und $B'_{k,1} = A'_{k,l}$ für alle $k \in \{1,\dots,n\}$. Damit folgt wie behauptet nach Satz 18.12 angewendet auf B

$$\det A = (-1)^{l-1}\det B = (-1)^{l-1}\sum_{k=1}^{n}(-1)^{k+1}\,b_{k,1}\,\det B'_{k,1} = \sum_{k=1}^{n}(-1)^{k+l}\,a_{k,l}\,\det A'_{k,l}.$$

(b) Dies ergibt sich mit Bemerkung 18.10 sofort aus (a). □

Beispiel 18.16 (Berechnung von Determinanten)**.** Die Entwicklung nach Laplace ist oft die geschickteste Art, die Determinante einer Matrix A konkret zu berechnen – insbesondere wenn man nach einer Spalte oder Zeile entwickeln kann, in der bereits viele Einträge gleich 0 sind, so dass die entsprechenden Terme in der Summe wegfallen. In der Praxis empfiehlt es sich daher, zunächst mit elementaren Spalten- oder Zeilenumformungen eine Spalte oder Zeile zu erzeugen, in der nur ein Eintrag ungleich Null ist, und dann nach dieser Spalte bzw. Zeile zu entwickeln. Beachte, dass die Determinante dabei nach Lemma 18.5 ...

- mit λ multipliziert wird, wenn wir eine Spalte oder Zeile mit λ multiplizieren; und
- unverändert bleibt, wenn wir ein Vielfaches einer Spalte bzw. Zeile zu einer anderen addieren.

Hier ist ein Beispiel, bei dem wir der Reihe nach die erste von der dritten Spalte subtrahieren, nach der dritten Spalte entwickeln, und noch einmal nach der zweiten Zeile entwickeln: Es ist

$$\det\begin{pmatrix}1&1&1&1\\1&0&1&0\\0&1&2&0\\0&1&0&3\end{pmatrix} = \det\begin{pmatrix}1&1&0&1\\1&0&0&0\\0&1&2&0\\0&1&0&3\end{pmatrix} = (-1)^{3+3}\cdot 2\cdot\det\begin{pmatrix}1&1&1\\1&0&0\\0&1&3\end{pmatrix} = 2\cdot(-1)^{2+1}\cdot 1\cdot\begin{pmatrix}1&1\\1&3\end{pmatrix} = -4.$$

Ein besonders einfacher Fall – der aber dennoch häufig vorkommt – sind die sogenannten Dreiecksmatrizen, bei denen oberhalb oder unterhalb der Diagonale nur Nullen stehen.

Definition 18.17 (Dreiecksmatrizen)**.** Eine quadratische Matrix $A = (a_{i,j})_{i,j} \in K^{n\times n}$ heißt **obere Dreiecksmatrix**, falls $a_{i,j} = 0$ für alle $i > j$ gilt, und **untere Dreiecksmatrix**, falls $a_{i,j} = 0$ für alle $i < j$ gilt. Obere bzw. untere Dreiecksmatrizen haben also die Form

$$\begin{pmatrix} a_{1,1} & & * \\ & \ddots & \\ 0 & & a_{n,n} \end{pmatrix} \quad \text{bzw.} \quad \begin{pmatrix} a_{1,1} & & 0 \\ & \ddots & \\ * & & a_{n,n} \end{pmatrix}.$$

Sind zusätzlich noch alle Einträge $a_{i,i}$ auf der Diagonale gleich Null, so heißt A **echte (obere bzw. untere) Dreiecksmatrix**.

Folgerung 18.18 (Determinante von Dreiecksmatrizen)**.** *Ist $A = (a_{i,j})_{i,j} \in K^{n\times n}$ eine (obere oder untere) Dreiecksmatrix, so ist ihre Determinante gleich dem Produkt ihrer Einträge auf der Diagonale*

$$\det A = a_{1,1} \cdot \dots \cdot a_{n,n}.$$

Beweis. Da untere Dreiecksmatrizen beim Transponieren in obere übergehen, reicht es nach Folgerung 18.9, die Aussage für obere Dreiecksmatrizen zu zeigen. Wir beweisen die Aussage in diesem Fall mit Induktion über n; der Fall $n = 1$ ist dabei trivial. Für $n > 1$ entwickeln wir $\det A$ gemäß Satz 18.15 nach der 1. Spalte: Da hier nur der erste Eintrag ungleich Null ist, ergibt sich sofort nach Induktionsvoraussetzung

$$\det A = (-1)^{1+1} a_{1,1} \det A'_{1,1} = a_{1,1} \cdot (a_{2,2} \cdot \dots \cdot a_{n,n}),$$

da auch $A'_{1,1} \in K^{(n-1)\times(n-1)}$ eine obere Dreiecksmatrix (mit Diagonaleinträgen $a_{2,2},\dots,a_{n,n}$) ist. □

Aufgabe 18.19.

(a) Berechne $\det(A^5)$ und $\det(5A)$ für die Matrix

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 2 & -1 & 3 \\ 4 & 1 & 8 \end{pmatrix} \in \mathbb{R}^{3\times 3}.$$

(b) Für $a_1,\dots,a_n \in K\backslash\{0\}$ zeige man

$$\det \begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & a_1 & 0 & \cdots & 0 \\ 1 & 0 & a_2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 1 & 0 & \cdots & 0 & a_n \end{pmatrix} = -\left(\prod_{i=1}^{n} a_i\right) \cdot \left(\sum_{i=1}^{n} \frac{1}{a_i}\right).$$

Wir wollen nun noch zwei Ergebnisse zu Determinanten beweisen, die mehr aus theoretischer als aus rechentechnischer Sicht interessant sind. Das erste betrifft inverse Matrizen: Ist A eine invertierbare Matrix, so haben wir in Algorithmus 14.12 ja bereits gesehen, wie man A^{-1} konkret numerisch berechnen kann. Mit Hilfe von Determinanten können wir nun auch eine explizite Formel für A^{-1} angeben – die allerdings den Nachteil hat, dass sie bei konkreten Berechnungen relativ aufwendig ist, weil für jeden Eintrag von A^{-1} eine eigene Determinante berechnet werden muss.

Satz 18.20 (Explizite Formel für die inverse Matrix)**.** *Es sei $A = (a_{i,j})_{i,j} \in K^{n\times n}$.*

(a) *Ist $C = (c_{i,j})_{i,j} \in K^{n\times n}$ die Matrix mit Einträgen*

$$c_{i,j} = (-1)^{i+j} \det A'_{j,i}$$

(beachte die Vertauschung von Spalten- und Zeilenindizes bei der Streichungsmatrix!), so ist $CA = AC = (\det A) \cdot E_n$.

(b) *Ist A invertierbar, so ist die inverse Matrix von A gegeben durch*

$$A^{-1} = \frac{1}{\det A} \cdot C = \left((-1)^{i+j} \frac{\det A'_{j,i}}{\det A} \right)_{i,j}$$

mit C wie in (a).

Beweis. Für alle $i,k = 1,\dots,n$ überprüfen wir den (i,k)-Eintrag des Matrixprodukts CA: Nach Definition 13.9 ist dies

$$\sum_{j=1}^{n} c_{i,j} a_{j,k} = \sum_{j=1}^{n} (-1)^{i+j} a_{j,k} \det A'_{j,i} \overset{18.15}{=} \det \begin{array}{c} \begin{array}{ccccc} & & \text{Spalte } i & & \\ & & \downarrow & & \end{array} \\ \begin{pmatrix} a_{1,1} & \cdots & a_{1,k} & \cdots & a_{1,n} \\ \vdots & & \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,k} & \cdots & a_{n,n} \end{pmatrix} \end{array},$$

wobei die zweite Gleichung genau die Entwicklung nach Spalte i ist, und die Matrix auf der rechten Seite aus A entsteht, indem die Einträge aus Spalte k auch in Spalte i geschrieben werden. Die Determinante dieser Matrix ist aber 0 für $i \neq k$ (da dann zwei gleiche Spalten existieren) und $\det A$ für $i = k$ (denn dann ist diese Matrix gleich A). Damit ist $CA = (\det A)\, E_n$. Analog zeigt man auch $AC = (\det A)\, E_n$ und damit Teil (a). Die Formel in (b) folgt daraus natürlich sofort mit Division durch $\det A$. □

Beispiel 18.21. Für eine 2×2-Matrix

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}$$

hat die Matrix C aus Satz 18.20 die Einträge

$$c_{1,1} = (-1)^{1+1} \det(a_{2,2}) = a_{2,2}, \qquad c_{1,2} = (-1)^{1+2} \det(a_{1,2}) = -a_{1,2},$$

und genauso $c_{2,1} = -a_{2,1}$ und $c_{2,2} = a_{1,1}$. Damit ist nach Satz 18.20 (b) im Fall einer invertierbaren Matrix also

$$A^{-1} = \frac{1}{\det A} \cdot \begin{pmatrix} a_{2,2} & -a_{1,2} \\ -a_{2,1} & a_{1,1} \end{pmatrix}.$$

Eine konkrete Anwendung von Satz 18.20 ergibt sich bei der Lösung linearer Gleichungssysteme: Sind $A \in \mathrm{GL}(n,K)$ eine invertierbare Matrix und $b \in K^n$, so wissen wir bereits, dass das Gleichungssystem $Ax = b$ für x die eindeutige Lösung $x = A^{-1}b$ hat. Da wir gerade mit Hilfe von Determinanten eine explizite Formel für die inverse Matrix A^{-1} gefunden haben, überrascht es nicht, dass wir auch für die Koordinaten dieses Lösungsvektors $x = A^{-1}b$ eine ähnliche explizite Formel herleiten können:

Satz 18.22 (Cramersche Regel). *Es seien $A \in \mathrm{GL}(n,K)$ und $b \in K^n$. Wir bezeichnen die Spalten von A mit $a_1,\dots,a_n \in K^n$. Dann ist die (nach Bemerkung 14.26 (b) eindeutige) Lösung des Gleichungssystems $Ax = b$ der Vektor $x \in K^n$ mit den Komponenten*

$$x_i = \frac{\det(\, a_1 \mid \cdots \mid a_{i-1} \mid b \mid a_{i+1} \mid \cdots \mid a_n \,)}{\det A}$$

für $i = 1,\dots,n$.

Beweis. Nach Satz 18.20 (b) und Definition 13.9 der Matrixmultiplikation ist x_i, also die i-te Komponente des Matrixprodukts $A^{-1}b$, gleich

$$x_i = \sum_{j=1}^{n} (-1)^{i+j} \frac{\det A'_{j,i}}{\det A} \cdot b_j \overset{18.15}{=} \frac{1}{\det A} \det(\, a_1 \mid \cdots \mid a_{i-1} \mid b \mid a_{i+1} \mid \cdots \mid a_n \,),$$

wobei die zweite Gleichheit die Entwicklung nach der i-ten Spalte ist. □

Beispiel 18.23. Wir wollen mit der Cramerschen Regel das lineare Gleichungssystem

$$\begin{aligned} x_1 + x_2 &= 2 \\ x_1 - 2x_2 &= -1 \end{aligned} \quad \text{lösen, also} \quad \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

Dies ist sehr einfach: Es ist

$$x_1 = \frac{\det \begin{pmatrix} 2 & 1 \\ -1 & -2 \end{pmatrix}}{\det \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}} = \frac{-3}{-3} = 1 \quad \text{und} \quad x_2 = \frac{\det \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}}{\det \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}} = \frac{-3}{-3} = 1.$$

Beachte jedoch, dass es für konkrete Gleichungssysteme mit mehr als zwei Variablen meistens sehr rechenaufwendig ist, die Cramersche Regel zu verwenden. Man wird diese Regel daher meistens nur für theoretische Überlegungen verwenden, in denen man eine konkrete Formel für die Lösung (und nicht nur ein Lösungsverfahren) braucht. Für numerische Berechnungen ist das Gauß-Verfahren in Satz 14.25 wesentlich effizienter.

Aufgabe 18.24. Es sei $A \in K^{n \times n}$ eine quadratische Matrix, die eine Blockgestalt der Form

$$A = \left(\begin{array}{c|c} B & * \\ \hline 0 & C \end{array} \right)$$

hat, wobei $B \in K^{m \times m}$ und $C \in K^{(n-m) \times (n-m)}$ selbst quadratische Matrizen sind. Zeige, dass dann $\det A = \det B \cdot \det C$ gilt.

Aufgabe 18.25. Es sei $A \in K^{n \times n}$ eine invertierbare Matrix mit ganzzahligen Einträgen.

Zeige, dass A^{-1} genau dann ebenfalls nur ganzzahlige Einträge hat, wenn $\det A = \pm 1$ gilt.

Literatur

[B] A. Beutelspacher, *Lineare Algebra*, Vieweg-Verlag (2003)
[BF] M. Barner, F. Flohr, *Analysis 1*, de Gruyter Lehrbuch (2000)
[E] H.-D. Ebbinghaus et al., *Zahlen*, Springer-Verlag (1988)
[Fi] G. Fischer, *Lineare Algebra*, Vieweg-Verlag (2002)
[Fo1] O. Forster, *Analysis 1*, Vieweg-Verlag (2011)
[Fo2] O. Forster, *Analysis 2*, Vieweg-Verlag (2010)
[G] A. Gathmann, *Algebraische Strukturen*, Vorlesungsskript TU Kaiserslautern (2023), https://agag-gathmann.math.rptu.de/ags
[GK] G.-M. Greuel, T. Keilen, *Lineare Algebra I und II*, Vorlesungsskript TU Kaiserslautern (1999), www.math.uni-tuebingen.de/~keilen/download/LectureNotes/linearealgebra.pdf
[J] K. Jänich, *Lineare Algebra*, Springer-Verlag (2010)
[K1] K. Königsberger, *Analysis 1*, Springer-Verlag (2003)
[K2] K. Königsberger, *Analysis 2*, Springer-Verlag (2003)
[M] T. Markwig, *Grundlagen der Mathematik*, Vorlesungsskript TU Kaiserslautern (2011), www.math.uni-tuebingen.de/~keilen/download/LectureNotes/grundlagen11.pdf

Index